Experiments and Analysis of Mitochondrial DNA Segregation, Yeast Polarization and RNA Polymerase Dynamics

Experiments and Analysis of Mitochondrial DNA Segregation, Yeast Polarization and RNA Polymerase Dynamics

a dissertation presented
by
Rishi Har Jajoo
to
The Committee on Higher Degrees in Systems Biology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Systems Biology

Harvard University
Cambridge, Massachusetts

February 2015

Thesis advisors: Johan Paulsson and Michael Springer Rishi Har Jajoo

Experiments and Analysis of Mitochondrial DNA Segregation, Yeast Polarization and RNA Polymerase Dynamics

Abstract

This thesis presents work done in three different topics in biology with a combination of experiments, mathematical modeling and quantitative data analysis.

All cellular materials are partitioned between daughters at cell division, but by different mechanisms and with different accuracy. Many macromolecules partition in proportion to cytoplasmic volume with significant errors at low numbers, while chromosomes accurately push half the copies to each cell regardless of volume. Little is known about organelles, but in *Schizosaccharomyces pombe* the mitochondria are pushed to the cell poles by the spindle, along with the chromosomes. Here we find that mitochondria spatially re-equilibrate just before division, and that the mitochondrial volume and DNA-containing nucleoids instead segregate in proportion to the cytoplasm inherited by each daughter. However, in contrast to other macromolecules, nucleoid partitioning errors are still strongly suppressed. This is ensured by control at two levels: mitochondrial volume is actively distributed throughout a cell in a process involving the microtubule associated protein Mmb1p, and nucleoids are spaced out in semi-regular arrays within mitochondria. The data also suggest that nucleoid replication control is passive and Poisson, and that low concentration noise is achieved by accurate segregation rather than corrective feedback control.

Patterning in organisms is both beautiful and functional, but we have limited

understanding of the general principles that cause patterns to emerge from isotropic initial conditions. Alan Turing proposed a minimal set of conditions that would allow reaction-diffusion systems to spontaneously form stable patterns. Since then many biological systems have been hypothesized to be of the Turing type, but with limited evidence that they actually rely on this mechanism. One of the conditions of Turing's theory is that reactions must have non-linear sensitivities (cooperativity) in their positive feedback cycles. We use a minimal model of yeast bud site selection to show that the amount of cooperativity achievable is naturally limited by the kinetic parameters achievable by proteins and the type of reaction mechanisms commonly used to achieve cooperativity. Further, higher levels of cooperativity necessarily cause pattern formation to slow down, likely causing a fitness defect for the organism. Then, because cooperativity is also used in many other biology systems, we show examples in TF-promoter binding and cooperative ligand binding where the trade-off between cooperativity speed of response is also present. These considerations bring into question how biological systems respond to these trade-offs. We then explore a conceptual model for yeast budding where an additional actin-based positive feedback loop uses a time-delay, not cooperativity, to ensure that yeast picks a single bud site.

RNA polymerase II (RNAP) pauses and backtracks during transcription elongation to regulate gene expression, control transcriptional efficiency and ensure transcriptional accuracy. After a pause, RNAP often backtracks and, in the presence of TFIIS, cleaves off the 3' end of nascent RNA, allowing productive transcription to restart. However, the precise determinants of RNAP pausing,

backtracking, and restarting remain unclear. Native elongating transcript sequencing (NET-seq) visualizes RNA polymerase pausing and backtracking with nucleotide resolution. Applying NET-seq to wild-type *S. cerevisiae* gives locations where cleavage occurs after backtracking. In contrast, NET-seq data from a strain where TFIIS is deleted reveal the primary points of pausing. Using these data, we show that RNAP pausing is largely controlled by the relative strength of the RNA:DNA and DNA:DNA hybrids in and downstream of the transcription bubble. In addition, backtracking is likely determined by the stacking interactions between the 3'-end of nascent RNA and Tyr769 of the Rpb2 subunit of RNAP. Though other factors beyond structural energetics of the transcription elongation complex certainly play a role in RNAP backtracking, this work suggests that the ubiquity of RNAP backtracking is largely controlled by sequence elements.

Contents

Author List

The following authors contributed to Chapter 2: Rishi Jajoo performed all experiments, most analysis and writing. Yoonseok Jung created the yeast strain YJ014. Matheus Viana and Susanne Rafelski created the MitoGraph software and extracted mitochondrial volumes and connectivity maps from images. Johan Paulsson and Michael Springer provided guidance and helped write the manuscript.
The following authors contributed to Chapter 3: Rishi Jajoo performed all simulations and analysis with guidance from Michael Springer and Johan Paulsson.
The following authors contributed to Chapter 4: Rishi Jajoo, Martin Lukacisin and Matthieu Landon each contributed equally to the analysis of NET-seq data. Rishi Jajoo made the strains with synonymous mutations, made the figures and wrote the chapter. Julie Crocker did the NET-seq experiments for synonymous mutations and the biological replicas of the original experiments in the lab of Stirling Churchman. Stirling Churchman provided guidance.

List of figures

Acknowledgments

Graduate school has been a rough ride, but I am better for it. I owe my successes to others. Chief among them are my parents who made me what I am and instilled in me the will to get things done. I have relied on my wife, Zeba Wunderlich, the most over these past years. She is my scientific and personal sounding board, and I have learned many lessons through her council. My extended family has also been an invaluable support network. Our cat, Mouse, has been a constant companion who has taught me about being playful.

Scientifically, I would like to thank Louis Staudt for a conversation 12 years ago that launched me in this direction. I owe a lot to Roy Kishony for taking me into his lab and paying me when I didn't know anything about biology.

Johan Paulsson is a man whose scientific and mathematical insight I will always admire and try to emulate. Michael Springer's wide interest in many parts of biology is refreshing and always helps to put my own work in perspective.

Many people in both the Paulsson and Springer labs also deserve my thanks: Shlomi Reuveni for being encouraging and showing me that you can be even-keeled even when facing uncertainty and for encouraging me to keep pushing on NET-seq. Andreas Hilfinger for never being too busy to entertain a crazy idea, do some math for me, talk sports or help with bike repairs. Nate Lord for being a mentor in the lab. His unending optimism still baffles me, and I hope to one day have his attitude. Laurent Potvin-Trottier for always being there to help me troubleshoot. Dirk Landgraf's inhuman work ethic taught me that if you keep on pushing to the unknown, something will come up that makes it all worth it. Billy Lau taught me

that just making one little thing work every day will eventually lead to a path out of here. Thomas Norman for his even keel and big picture thinking. Greg Foley for making sure that I realized that there is a larger world outside the lab. Dann Huh for starting the partitioning project, and offering advice for surviving in the lab. Jue Wang for being a great hiking buddy. Meriem El-Karoui for teaching me sterile technique when she saw me struggling. Jennifer Waters for teaching me all I know about microscopy. I would also like to thank all the members of the Paulsson and Springer labs who have been great labmates: Bo, Ilan, Christine, Sarah, Yoni, Renan, Raul, Shay, Per, Scott, Charles, Ghee Chuan, Yoon, Burak, Anna, Emanuele, and Jeanne.

I am fortunate to have collaborated with Susanne Rafelski and Matheus Viana. Their kindness, decency and excitement is something I treasure.

The most rewarding part of my PhD was spent working with Matthieu Landon and Martin Lukacisin. The easy exchange of ideas and palpable excitement in finding new patterns in the data is something that I hope to recapture in the future. I would also like to thank Stirling Churchman for being a sounding board in our work on NET-seq.

I would also like to thank members of the DePace Lab who have been awesome in their social support. I would especially like to thank Meghan Bragdon, Max Staller and Ben Vincent were always there to hear me out and make me laugh.

There is no way I could have cleared all the hurdles of graduate school without support from Samantha Reed. Her constantly positive attitude and unflappability is something I try to emulate. I would also like to thank Becky Ward for incisive advice on science and writing. Barb Grant, Kathy Buhl, Cris Roberts and the rest of the lab-ops, glass wash, and administrative staff have made the trains run on time and are also just great people. My DAC committee members–Fred Winston, Vamsi Mootha and Tim Mitchison– have been a great source of practical advice for publishing and graduating.

I would like to thank my gah buddies Bailes and Emily Brown, Michelle Wu, Conor Pewarski, Ed and Tizzie Likovich, Geoffrey Peterson, Andrew Dranginis, David Staudt, Crystal and Jared Bevis, and Rolando Flores for helping me

remember to have fun. I would also like to thank Glynis Ritchie for her friendship and lawyering skills.

Even though I was not her lab, I would like to thank Angela DePace because her focus on people has been a beacon for how to do science while not being consumed by its ups and downs. Through Zeba, Max, Ben and Meghan I have learned that the goal of a lab is to help people get to where they want to go and that there is no shame in doing science in accordance with your values.

1

Introduction

This thesis presents work done in three disparate areas of quantitative biology. My main project has been to quantify the accuracy of segregation and replication of mitochondria and mitochondrial DNA in the fission yeast *Schizosaccharomyces pombe*. I also worked to create a mathematical understanding of how budding yeast can reliably choose a single site from which to bud. In addition, I investigated how

functional sequencing data could be used to create a biophysical model of RNA polymerase dynamics.

The principle of pairing chromosomes and pushing them apart at cell division is often painted as a beautiful ballet, and I am interested in how other cellular components dance. The problem of cell division is often reduced to that of chromosome segregation, but each and every component must be segregated to daughter cells. Huh and Paulsson showed the importance of segregation in controlling heterogeneity between cells and how difficult it is to reduce this noise [1, 2]. One key insight from their studies was that partitioning control of multilevel components, such as organelles, has to be done at each level in order for control to be effective at the bottom level. They also showed how different partitioning mechanisms could have different statistical error signatures, but that many radically different segregation mechanisms could also have very similar segregation statistics [1, 2]. Given the number of mechanisms that are employed to make chromosomes pair off and push to opposite ends of the cell, the other components of a cell that are likely to also have interesting partitioning mechanisms.

We chose to study the segregation of mitochondria and mitochondrial DNA (mtDNA) because of their importance in cell physiology and their intricate multilayered structure: multiple mtDNA genomes are packaged with protein into bundles called nucleoids which are present in tens to hundreds of copies within the mitochondria, which is itself undergoing constant fission and fusion. The mitochondrial body is also interesting because in *S. pombe* are pushed to the end of the cells with the nucleus during cell division, suggesting that yeast might have

active mechanisms to segregate mitochondria [3]. Similarly, in budding yeast, Rafelski et al. have found that mother and daughter cells inherit different concentrations of mitochondria through an active mechanism that preserves newly born cells at the expense of the mother [4]. We were also able to collaborate with the Rafelski group to use their image analysis technology to calculate the absolute volume of mitochondria in dividing cells.

Previous efforts to study the process of mitochondrial inheritance focused on qualitative features such as production of petite yeast (those without functional mtDNA) and morphological descriptions of mitochondria to identify fission and fusion proteins [5–8]. Most quantitative studies of mtDNA partitioning and replication have been theoretical or *in vitro* [9–11]. One notable exception is in scorpion sperm where nearly equal numbers of mitochondrial bodies are divided between cells [12]. Attempts to monitor mtDNA in single cells have been difficult because of interference from competing signals of the vast excess of nuclear DNA [13, 14]. I developed assays to overcome these challenges by creating one new method (Sybr Green I staining) and using an established method more quantitatively (EdU staining [15]).

Replication of mtDNA has not been studied at the single cell level. It is known that each mtDNA molecule has the same chance of replicating [16], but no replication control mechanisms are known. We were especially interested in control of mtDNA replication because it is a self-patterning molecule that could undergo runaway replication without other mechanisms of control and because it is likely needed in proportion to the cell volume.

The third chapter of this thesis explores the effects of cooperativity in spontaneous pattern formation in biological systems. To understand how spontaneous pattern formation is achieved, Alan Turing proposed the now famous strategy requiring stochastic initiation followed by local non-linear positive feedback and global negative feedback [17, 18]. However, Turing's theory only indicates when stable patterns can exist, not if this process can meet the other objectives under realistic biophysical constraints. Very few studies have worked to understand the general trade-offs that might occur when implementing a Turing model [19, 20]. We explored some of the complications that arise from one of the key assumptions of Turing-type pattern formation: cooperative (super-linear) feedback which has often been assumed but not explicitly tested.

To understand the basic consequences of cooperativity in spontaneous pattern formation, we chose to study the "hydrogen atom" of spontaneous pattern formation: how the yeast *Saccharomyces cerevisiae* chooses a single bud site from an isotropic cell membrane. Specifically we thought about the criteria that selection might use to shape the mechanisms of bud site selection: quickness in polarizing, uniqueness in polarization site and stability of site once it is formed. We then explored how cooperativity changed the ability for a polarity system to meet these criteria and propose how a second feedback loop that works on a time-delay principle could allow for a non-cooperative mechanism of polarity. This allowed us to also observe a more general trade-off between cooperativity and speed of reactions that could apply to other biological systems as well.

The fourth chapter of this thesis builds a biophysical model of RNA polymerase

(RNAP) pausing and backtracking by mining functional sequencing data. RNA polymerase dynamics have long been studied as molecular machines at single-base resolution with single molecule methods [21–23], or at the whole genome level [24, 25], but these two approaches could not be combined because single molecule methods lacked the sequence diversity and *in vivo* context and genomic methods lacked single-base resolution. The current hypothesis for RNAP pausing is that nucleosomes are a major barrier to transcriptional elongation.

Nascent Elongating Transcript sequencing (NET-seq) [26] allows genome-wide RNA polymerase density at single base resolution and we reasoned that this data would allow us to detect the sequences that caused RNAP pausing and backtracking due to simple interactions in the transcription bubble. Such interactions have been previously investigated in theoretical studies [27, 28]. Most studies used mutated polymerase residues as a way to understand interactions between sequence and RNAP [29–32], but we realized that NET-seq could be used understand how native RNAP responds to many different sequence contexts. Then by turning RNAP density into a quantitative prediction game, we were able to then test huge numbers of sequence-based hypotheses against the data and settle on to the most plausible ones.

Although these projects have little in common from a biological perspective, they are cases where theory and experiments combine in a productive way to help solve interesting questions about our world.

2

Accurate concentration control of mitochondria and nucleoids

Mitochondria are cytoplasmic organelles present in most eukaryotes. They generate much of the chemical energy of cells, and play key roles in signalling, apoptosis and disease [33–35]. However, little is known about how their volumes or their

DNA-containing nucleoids are controlled during the cell cycle or at cell division. From first principles, accurate partitioning of nucleoids seems particularly challenging because segregation must be precisely controlled at two levels: even if nucleoids segregate accurately within mitochondria, errors between cells can still be large unless the mitochondrial volume also segregates accurately [1, 36]. Feedback control of production is in turn challenging due to delays or noisy signaling in the feedback loop – issues that turn out to be particularly problematic for components in low numbers [37, 38].

We therefore set out to determine the accuracy of noise control in mitochondria and mitochondrial nucleoids in the well-studied, symmetrically dividing fission yeast *Schizosaccharomyces pombe*, where we developed methods to count nucleoids and precisely measure the volume of mitochondria in single cells.

2.1 Volumetric partitioning of mitochondrial and nucleoids

We first used time-lapse imaging of mitochondrial matrix-targeted mCherry to quantify spatial dynamics during the cell cycle and after cell division. This confirmed reports that mitochondria are pushed to the poles of the cell by the mitotic spindle [39, 40] before cell division (Fig. 2.1 A), in roughly equal amounts regardless of where the cell eventually divides. Surprisingly, however, we further found that in the last 15% of the cell cycle – after chromosome segregation but before cytokinesis – mitochondria escape from the poles and spatially re-equilibrate throughout the length of the cell, in virtually every cell observed both for wild type cells and for cells that divide asymmetrically (*pom1*Δ, Fig. 2.5 and 2.1 B,C). This

suggests that partitioning occurs by a qualitatively different principle, conserving concentrations rather than numbers when cell division is not perfectly symmetrical, as further confirmed by measuring both the mitochondrial (via the MitoGraph technique [41]) and cytoplasmic volumes in each daughter cell (Fig. 2.1 D).

Since the observed migration of the mitochondria to the poles is so suggestive of an active segregation mechanism, and even coincides with the segregation of the nuclear DNA, we considered whether mitochondrial nucleoids might still be segregated via this mechanism. Specifically, the nucleoids could remain in the cell half where they are originally placed by the spindle, even after the majority of the mitochondrial volume re-equilibrated. Counting mitochondrial nucleoids was not previously possible because of the interfering signal from nuclear DNA and thus virtually nothing is known about how they segregate. We developed a fluorescent microscopy method using the dye Sybr Green I (SGI, Molecular Probes) that does not significantly stain nuclear DNA, and cross-validated the results with two other single cell methods – one other dye that can be used simultaneously in the same cells as SGI, and one method based on "click" chemistry that can only be used in specific mutants [42]. This showed that SGI indeed stains almost all nucleoids in the cell (see methods). The caveat of this method is that the necessary washing procedure perturbs the arrangement of mitochondria in the cell; thus it is possible to count the number of nucleoids in each cell but not to determine their spatial positions. We therefore used it to test whether the number of nucleoids in each daughter cell was dependent on how the cytoplasm is divided, i.e. whether they segregate by number, like chromosomes (Fig. 2.1 D, "Equal number" line), or by concentration, like RNA

or protein molecules (Fig. 2.1 D, "Equal concentration" line). The data confirmed that nucleoids segregate in proportion to cell volume, showing a near-perfect match between the fraction of cytoplasm and the fraction of nucleoids.

2.2 Better than binomial segregation of mitochondria and nucleoids

Segregation in proportion to the available cytoplasmic volume in each cell (Fig. 2.1 D) may conserve concentrations in the two daughters *on average*, but this typically comes at the cost of large relative partitioning errors at low numbers [1]. Given the relatively low number of nucleoids per cell ($\sim$ 30 per dividing cell), this should create significant noise in nucleoid concentrations ($\sim$15%) even if the mitochondrial volume partitioned perfectly. Surprisingly, however, counting the number of nucleoids and quantifying the mitochondrial volume that segregated to each daughter cell at division revealed that this is not the case: daughters of approximately equal sizes inherit very similar numbers of nucleoids – far more similar than expected from binomial partitioning (Fig. 2.2A) and possibly even close to perfect when correcting for experimental inaccuracies (see SI). Strikingly, even asymmetrically dividing cells (*pom1*Δ) (Fig.2.2B, center) or cells that divide at half the normal length (*wee1*Δ) (Fig.2.2B, right) partition nucleoids proportional to cell volume with high accuracy.

Since mitochondria form a semi-continuous mesh, as opposed to the discrete and stereotyped units of most textbook cartoons, passive mechanisms could explain the observation that the mitochondrial volume closely tracks the available cytoplasmic volume. The sub-binomial error of nucleoids within mitochondria, however, requires

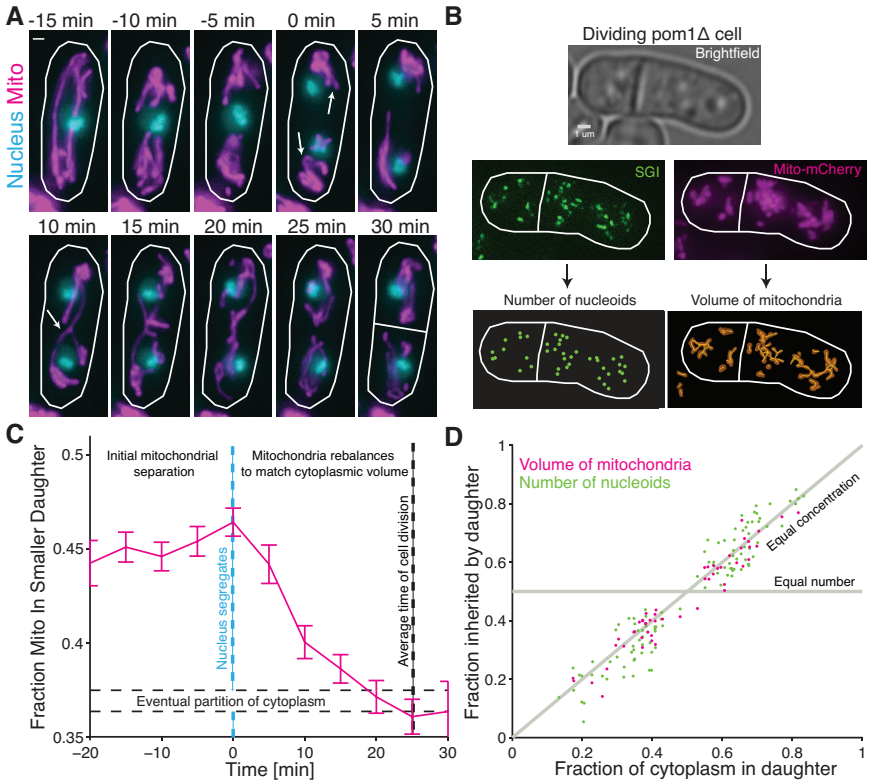

Figure 2.1: Mitochondria and nucleoids spatially re-equilibrate after nuclear segregation to match cytoplasmic division. **(A)** Timelapse images of a dividing wild-type cell (RJP041) in brightfield and with mitochondrial matrix-targeted (magenta) and GFP labeled nuclei (cyan). Arrows at 0 min indicate initial separation of mitochondria at the same time as nuclei and then reformation of a continuous mitochondrial network at 10 min but before division at 30 min. **(B)** For individual dividing *pom1*Δ cells (RJP025), the number of nucleoids in each daughter cell was determined with semi-automated spot detection of SGI signal and the mitochondrial volume in each cell half was determined using the surface-based volume reconstruction from the MitoGraph software package [41]. **(C)** Fraction of mitochondria in the part of the cell that will eventually become the smaller daughter as a function of time for *pom1*Δ cells (RJP042). Images of cells as in (A) were taken every 5 minutes and aligned at 0 minutes (blue dashed line) when nuclei first separate. Average cell division occurred 25 minutes after nuclear separation with approximately 37% of cytoplasmic volume inherited by smaller daughter (dashed horizontal lines). Error bars indicate standard error of the mean. (n=91 cells) (**D**) For *pom1*Δ cells, the fraction of nucleoids and of mitochondrial volume in each daughter cell as computed in (B) are plotted as a function of the fraction of cytoplasmic volume in each daughter cell. Gray lines indicate expectations for equal concentrations or equal numbers of mitochondria or nucleoids segregating to each daughter cell. (n=102 cells). All scale bars indicate 1 um.

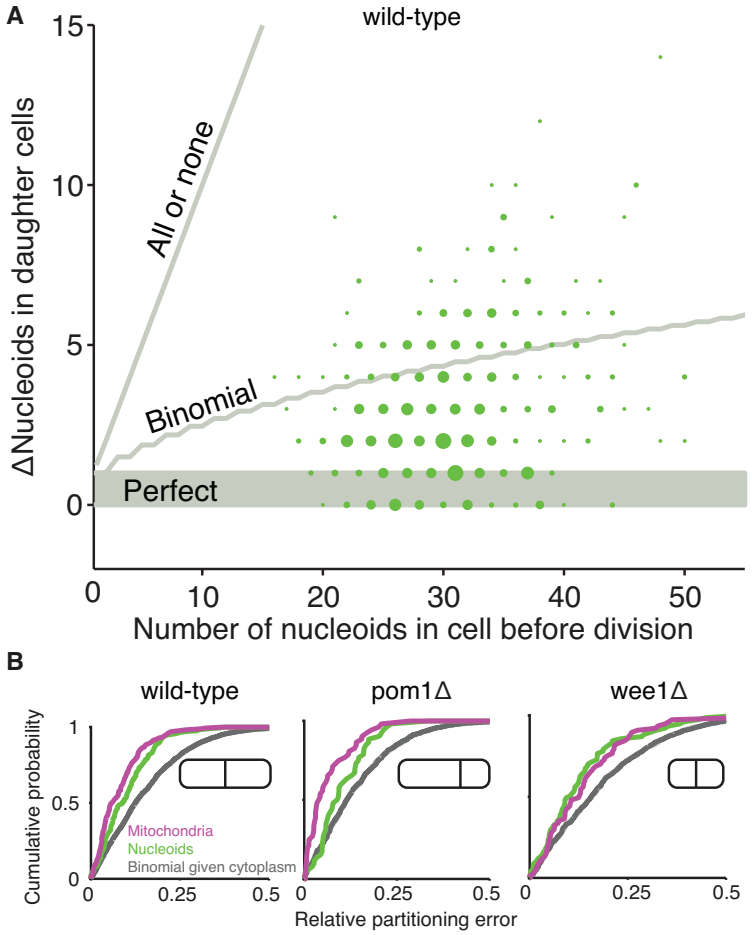

Figure 2.2: Mitochondria and nucleoids are partitioned to daughter cells more accurately than expected by passive (binomial) partitioning. (**A**) The absolute difference between the number of nucleoids segregated to each wild-type (RJP005 and DH60) daughter cell is plotted against the total number of segregating nucleoids. Area of plotted point is proportional to the number of samples observed with those counts. Gray lines indicate models of perfect, binomial (passive) and all-or-none segregation. Most cells partition better than the binomial expectation and many cells partition with perfect fidelity. Binomial segregation predictions were calculated by expecting that each nucleoid had a 50% chance of being inherited by each daughter (n=420 cells) (**B**) Cumulative distribution functions (CDFs) of the relative partitioning errors of mitochondrial volume (magenta) and nucleoid partitioning (green) compared to cytoplasmic partitioning for wild-type (RJP005, n=168 cells), *pom1*Δ (RJP025, n=52 cells), and *wee1*Δ (RJP029, n=77 cells). Mitochondria and nucleoids were more accurately partitioned than the binomial model would predict as CDFs are to the left of the binomial expectation (t-test, Nucleoids: wild-type: $p < 10^{-15}$, *pom1*Δ: $p < 10^{-3}$, and *wee1*Δ: $p < 10^{-3}$); mitochondria: wild-type: $p < 10^{-27}$, *pom1*Δ: $p < 10^{-11}$, and *wee1*Δ: $p < 10^{-2}$).

some sort of active spatial control. Since microtubules are known to be involved in mitochondrial segregation, we examined *mmb1*Δ mutants where mitochondria are not able to attach to microtubule filaments [43]. This led to highly inaccurate segregation of mitochondrial volume (Fig.2.3 A). Strikingly, however, the nucleoids still segregated with sub-binomial errors after the errors in the mitochondrial volume had been corrected for by conditioning the data. Thus the volumetric partitioning of mitochondria is active rather than passive just as in budding yeast [4], perhaps because mitochondria are too large and filamentous [39, 43, 44] to rely on passive processes. A second, independent, active mechanism is thus required to ensure that mitochondrial segments of the same size contain the same number of nucleoids.

2.2.1 Regular spacing accurately couples nucleoid segregation to mitochondrial segregation

To identify the principle by which mitochondria ensure a non-random fraction of nucleoids in each daughter – regardless of the position of the septum – we developed a method to accurately quantify the spatial locations of nucleoids. We used a strain of *S. pombe* that expressed a mitochondrial matrix-targeted mCherry and also could incorporate 5-ethynyl-2'-deoxyuridine (EdU) into nucleoids [42] which could be later visualized through a click-chemistry reaction [45]. The caveat of this method is that it requires cell-cycle arrested cells which are then fixed. However, unlike the SGI assay above, this method allows us to visualize native mitochondrial morphology and detect nucleoids in interphase cells (see methods). Mapping both the mitochondrial network and the position of nucleoids of cells (Fig.2.3 B) using

MitoGraph and the nucleoid image analysis method used for SGI staining, showed that nucleoids are regularly spaced out within the mitochondria (Fig.2.3 C). Previous observations in human cells have also identified linear sections of mitochondria where nucleoids appear to be regularly spaced out [46, 47]. Our results show that nucleoids are in fact arranged extremely regularly throughout the mitochondrial volume and that this explains the low segregation errors (Fig.2.3 C inset), similar to a phenomenon observed for carboxysomes in cyanobacteria [48]. The molecular mechanisms underlying the even spacing are not known (Fig.2.3 D), but recent observations in budding yeast suggest that it could be related to a coupling of nucleoids to the endoplasmic reticulum (ER). The ER delivers membrane to the mitochondria to allow the mitochondria to grow [49]. Murley et al. showed that contact between the ER and mitochondria occurs preferentially near replicating mtDNA [50]. This suggests that membrane could be preferentially inserted between newly replicated mitochondrial genomes, helping to separate and evenly distribute them along the mitochondria.

2.3 Nucleoid replication

The methods above can also be used to indirectly evaluate the accuracy of nucleoid replication control during the cell cycle. Specifically, we measured the distribution of the number of nucleoids in cells that had just divided, capturing both the beginning and end of the cell cycle (Fig.2.4 A), both for wildtype cells and for *mmb1*Δ, *pom1*Δ and *wee1*Δ mutants where segregation errors contribute differently to the total noise (Fig.2.4 B) but where production control should be intact.

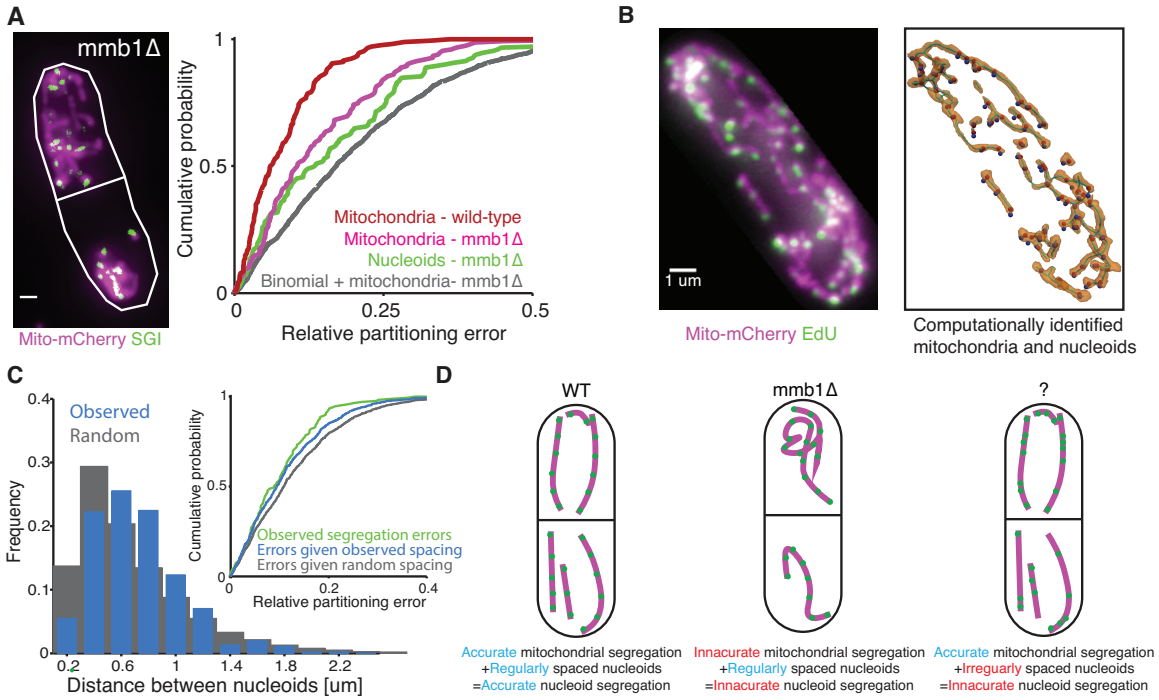

Figure 2.3: Accurate segregation of nucleoids requires both accurate mitochondrial segregation and regular spacing of nucleoids in the mitochondrial network. (**A**) CDFs of partitioning errors for mitochondrial volume and nucleoid segregation (as in Fig. 2.2B) in *mmb1*Δ cells (RJP016). Wild-type mitochondrial errors are shown for reference. Mutant *mmb1*Δ cells had higher errors than wild-type cells but still partitioned nucleoids more accurately than a binomial model conditioning on mitochondrial segregation. **(B)** A fixed cell with mito-mCherry (magenta) with EdU-labeled nucleoids (green) (RJP028) and a computational representation. **(C)** A histogram of the actual distances between neighbor nucleoids (blue, mean: 0.72 um, s.d.: 0.33 um) and randomly redistributed nucleoids within the mitochondrial network (gray, mean: 0.68 um, s.d.: 0.43 um). (1,871 nucleoids in 24 cells). Inset: CDF of relative nucleoid segregation errors for wild-type cells from Fig. 2.2B (green) and models of spacing from the observed interdistances (blue) and random spacing (gray). The random spacing model results in a significantly higher error than the observed wild-type segregation (t test, $p < 10^{-6}$). However the observed spacing does not fully recapitulate the low errors of the observed wild-type segregation (t test, p=0.02). 100 distributions each for 168 wild-type cells were performed assuming linear mitochondria (see methods). (**D**) In WT cells, mitochondria evenly arranged throughout the cell just before cell division and nucleoids inherit this accuracy by spacing out evenly in the mitochondrial network. In *mmb1*Δ cells, mitochondria are inaccurately partitioned, and nucleoids are also innacruately partitioned because they are equally spaced out in the mitochondria. Disrupting the spacing of nucleoids within the mitochondria through an unknown mutation would leave the mitochondrial segregation unchanged but would cause inaccurate nucleoid partitioning

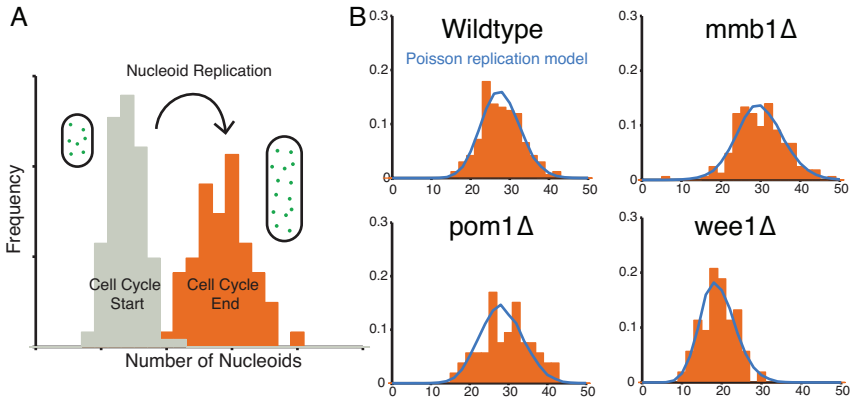

Figure 2.4: Nucleoid replication is independent of the number of nucleoids and free of extrinsic noise. (**A**) Schematic of data compiled from each experiment at the aggregate level. In dividing cells, the concentration of nucleoids was measured for both the newly born (grey) and the just divided cells (orange) and then normalized to the average volume of each population. (**B**) A Poisson process with no free parameters was simulated until the number of nucleoids doubled and the resulting distribution is plotted in blue against the actual data in orange for each of the strains reported. (n=200 simulations for each dataset)

Strikingly, all these strains approximately add a Poisson distributed number of nucleoids during the cell cycle (Fig.2.4 B). The Poisson distribution is expected in the absence of self-control, i.e., when new nucleoids are added independently of the number of nucleoids currently in the cell. This makes sense in the light of theoretical works demonstrating the caveats of feedback control – for components present in low numbers, noisy signaling, lags and delays, or limited cooperativity can greatly reduce the accuracy of control, and passive mechanisms may then perform even better than active ones [37, 38]. In fact, for passive mechanisms, the Poisson distribution is remarkably narrow since other sources of cell-to-cell variation in the production pathways could randomize levels further [51, 52].

2.4 Discussion

Our results show that the segregation of both mitochondrial volume and mitochondrial nucleoids are accurately controlled in single cells, by methods that conserve concentrations rather than total numbers, and that rely more on accurate partitioning than on self-control of production. This makes them qualitatively different from both macromolecules and chromosomes: chromosomes use tightly controlled replication and partitioning mechanisms that ensure equal numbers to each daughter, while cytoplasmic macromolecules segregate by volume but with substantial errors both from gene expression bursts and from segregation. The accurate concentration control of mitochondria may reflect the fact that mitochondria are a key source of metabolite, lipid and energy production for the cell, and are therefore needed in tight proportion to the cell size [4].

The fact that nucleoid segregation is better than binomial and nucleoid replication is a Poisson process has implications for how cells control concentrations. Heterogeneity between cells is usually generated at cell division and then corrected by replication during the cell cycle. However, in this case, it is "sloppy" replication control that is corrected by accurate segregation, a reversal of the usual paradigm.

2.5 Materials and methods

2.5.1 Nucleoid counting assay

We developed a novel assay for counting individual mitochondrial nucleoids in single dividing cells: live *S. pombe* stained by Sybr Green I (SGI) showed clear puncta resembling mitochondrial nucleoids, had no detectable signal in the nucleus and had minimal autofluorescence (Fig. 2.6 A). These SGI puncta co-localized with a mitochondrial matrix-targted mCherry fluorophore, and were not observed in cells lacking mtDNA (ρ^o) (Fig. 2.7 B). To verify that SGI puncta were nucleoids, we co-stained cells with a previously used nucleoid stain, DAPI [14] and saw a close correspondence between individual nucleoids identified by each method. Using a semi-automated image analysis method to count the number of nucleoids stained (see below), we also found that the average number of nucleoids detected by each method was nearly identical (Fig. 2.6B). This procedure produces two estimates for each individual cell, and we found that, on average, the low estimate was approximately 92% of the high estimate. To account for the fact that both methods then slightly undercount nucleoids, we performed a more rigorous statistical analysis showing that each nucleoid is independently detected with an efficiency of about 84% with either stain (Fig. 2.8).

Estimation of detection efficiency from DAPI and SGI staining

To estimate the staining efficiency of DAPI and SGI, each cell as assumed to have some number of nucleoids, N, and each stain was assumed to detect individual

nucleoids with a probability, p. This probability was assumed to be the same for each stain as the correlation coefficient between them was 1.0. For each cell, the average number of spots detected via DAPI and SGI was used as an estimate for Np. All cells with the same number of average detected nucleoids were grouped and the individual samples were treated as binomial process with N trials, each with p chance of success. The variance of this process is $Np(1-p)$ and p was computed by assuming that the average number of detected nucleoids for the group was Np. The result from each of these groups of cells was averaged with weight proportional to the number of samples and resulted in the estimated 84% detection efficiency.

Estimation of detection efficiency from EdU staining

To further validate SGI staining we compared it to cells where nucleoids were labeled with the nucleotide analog 5-ethynyl-2'-deoxyuridine (EdU), followed by "click" ligation of Alexa-488 azide (see below). EdU labels nucleoids reliably [42], but only in fixed cells after a cell cycle arrest (Fig. 2.9 A) and thus can only be used for assay validation purposes. In a matched population of cells, SGI stained approximately 84% of nucleoids stained by EdU (Fig. 2.9 B), agreeing with our previous estimate of labeling efficiency. SGI staining is therefore an accurate method for quantifying nucleoid numbers in single cells, but has two significant limitations: the experimental procedure disrupts the native localization of mitochondria and nucleoids, and nucleoids do not remain labeled if cells resume growth. Therefore, this assay cannot be used to monitor nucleoid position or number over time or in space. However, it does allow us to count the number of nucleoids in each daughter

of dividing cells, and when combined with judiciously chosen cell division mutants, can be used to identify key properties of the nucleoids segregation process.

Correction for detection efficiency

The observed segregation errors indicate that nucleoids partition to daughter cells more accurately than a binomial model (Figure 2) However, we wanted to ensure that this result held true even though we detected only $\sim 80-90\%$ of nucleoids. Huh and Paulsson showed that unbiased incomplete detection would make the data look more like binomial segregation than it actually is [1]. Because we observed better than binomial segregation without any corrections, this would imply nucleoids are actually even more accurately segregated when the imperfect detection is accounted for.

To quantitatively correct for detection error of nucleoids, we use equation S81 of Huh and Paulsson [1]. If l and r represent the number of nucleoids detected in the left and right cells respectively, L and R are the number actually in the left and right cells, and u is the efficiency of detection, then the conversion between the average squared error observed and the average squared error actually present is

$$\frac{1}{u^2}\Big[\langle (l-r)^2\rangle + (u-1)\langle l+r\rangle\Big] = \langle (L-R)^2\rangle \tag{2.1}$$

The result of correcting the observed data in this way is shown in Figure 2.10 shows that the nucleoid segregation is likely even more accurate than we observed. We plotted the uncorrected data in Figure 2 of the main as a conservative estimate of the error in segregation.

2.5.2 Correspondence with previous measurements

Recently divided cells showed an average of 15.4 nucleoids. Using previous estimates of the number of mtDNA genomes in newly divided *S. pombe* cells (100 genomes) [53], and the computed detection efficiency of SGI staining (84%), this implies that each nucleoid contains roughly 5 genomes on average, within the range reported in other organisms [47].

2.5.3 Staining procedure

Sybr Green I and DAPI staining

All cultures were grown in YES media with 4% glucose at 30 C before staining experiments. Overnight cultures were diluted approximately 100 to 1000-fold, allowed at least 2 doublings and grown to $OD_{600} \sim 0.1$. Approximately 1mL of culture was spun at 3000g for 1 minute and washed with 1x PBS. Volume was brought up to 100 uL (PBS) and 0.5 uL of Sybr Green I (Life Technologies, S-7567) (undiluted from stock) and/or DAPI (Sigma-Aldrich, D9542) (5 ug/uL) was added and contents were gently mixed. Cells were then immediately spun (as above) and washed with PBS twice before imaging.

EdU staining

For EdU incorporation into strains yFS284 and RJP028, strains were grown from overnight cultures at 27°C and then diluted approximately 100x and grown until $OD_{600} \sim 0.1$ at 27°C ($\sim$2 doublings). Cells were then spun down (as above) and

resuspended in media at 37°C and incubated at 37°C for 30 minutes before EdU was added to 10 uM final concentration. Cells were then further grown at 37°C for 90 minutes and either fixed with ethanol immediately (see below) or 37% formaldehyde was added to cultures to 3.7% final concentration and shaken at 37°C for 30 minutes. Cells were then spun down (as above) and washed in PBS twice.

Ethanol fixation, either alone or after formaldehyde fixation, was performed by spinning down cells (as above) and replacing PBS with approximately 1 mL 70% ethanol. Cells were vortexed until pellet dissolved. Cells were then spun down (as above) and washed twice with PBS, taking care to dissolve the pellet each time. EdU was ligated to Alexa-488 azide (Life Technologies) essentially as described by Hua and Kearsey [15] except that PBS buffer was used in place of TBS buffer. Note that formaldeyhyde fixation was required to preserve mCherry signal and ethanol fixation was required to perform "click" reaction.

Agarose pad for imaging was created as described previously [54] except that two layers of frame seals (Biorad, SLF-1201) were placed on microscope slide to contain slab. Agarose pad was made from 2% low fluoresence agarose (Biorad, Catalog #161-3100) in PBS (for spot counting) or YES media (for time-lapse movies). Cells were allowed to dry on agar pad for approximately 10 minute before a No 1.5 coverslip (VWR) was placed on top.

2.5.4 Microscopy

Microscopy was performed on a Nikon Eclipse Ti inverted microscope equipped with an Orca R2 (Hamamatsu) camera, a 100x Plan Apo oil objective (NA 1.40, Nikon),

an automated stage (H117, Prior Scientific), and a Lumen 200 Pro metal arc lamp illumination system (Prior Scientific). Image acquisition was performed using μManager or μManager controlled by custom Matlab scripts. For time-lapse imaging the microscope was encased in a custom-built incubator maintained at 30°C throughout the experiment. The following filter cubes (Semrock) were used for image acquisition: DAPI (excitation 390/40, dichroic 405, emission 452/45), GFP (excitation 472/30, dichroic 495, emission 520/35), YFP (excitation 500/24, dichroic 520, emission 542/27) and mCherry (excitation 562/40, dichroic 593, emission 641/75).

For spot detection, images for mCherry (mCherry filter) and SGI (YFP filter) and DAPI (DAPI filter) were taken in Z-stacks with 0.25 um spacing in a 6 um range with 200-300 ms exposure time. Brightfield images were taken at a single Z position that clearly showed the cell wall and any septum with 100ms exposure.

For timecourses, images were taken in mCherry and GFP channels in Z-stacks with 1.0 um spacing over 4 or 5 um range and 50 ms or 40 ms exposure in each channel. The percent of mitochondria in the smaller daughter was determined as the fraction of the total fluorescence in the mCherry channel that was in the smaller daughter. The fluorescence for each cell half was determined by first subtracting background and then adding pixel values from all Z-planes. Cells with obvious mitochondrial photodamage (permanently balled, fragmented or aggregated mitochondria) were excluded from analysis.

2.5.5 Image analysis

To count nucleoids in dividing cells, dividing cells were first identified by their straight and complete septa in bright field images and their outlines and septa traced using the microbeTracker [55] software for Matlab. Cells with incomplete or rounded septa were excluded from analysis as they had not divided or had been divided for more than 15 minutes (D. Huh, personal communication).

To identify nucleoids, custom Matlab (Mathworks, version 2013a) code was written to sharpen spots using a 3-dimensional unsharp mask (blurring ellipsoid was 3 pixels by 3 pixels by 1 pixel (xyz)). The sharpened image was then thresholded on the first local maximum of the entropy distribution of pixel intensities. The Matlab morphological commands bwmorph:clean, bwmorph:fill and imopen (with disk of radius 1) were then used on each slice. Connected areas in the 3D image (using the 6 neighbor definition) with fewer than 20 pixels were then eliminated. All other connected pixels were provisionally considered nucleoids. These identifications were confirmed by viewing the sharpened maximum projection of the image and manually correcting any missed or erroneously detected spots using the spotFinderM feature of microbeTracker. As spots had a broad range of intensities, the maximum projection of a a log transformed version of the image was also checked for any missed spots and manually corrected. All code available upon request.

2.5.6 Data analysis

Cell volume estimation

To estimate the available cytoplasmic volume in each cell half at division the nucleus was estimated to occupy 11% of the cytoplasm before division and equal sized nuclei are transported to each cell half, regardless of the sizes of the cells. Previous data report between 8% [56] and 12% [57] of the cell is composed of the nucleus. Using an average of these two estimates and that 10% of the cell is composed of cell wall [57], then roughly 11% of the cytoplasmic volume is occupied by the nucleus.

Distances between nucleoids

To calculate the distances between two nucleoids in the mitochondrial network, first, nucleoids were identified as previously described and computationally overlayed onto the network. Nucleoids were assigned to the point on the network with the nearest Euclidean distance. For this analysis mitochondria were considered 1-dimensional objects and two nucleoids were considered on the same edge if they were connected by unbranched mitochondria. Distances between nucleoids on unconnected mitochondria were not computed. Distances between nucleoids where the mitochondria between them was branched were also not computed. Only segments greater than 1um were considered as these are unlikely to be spurious artifacts of the image analysis. As the diffraction limit for the Alexa-488 was was 0.227 um, any nucleoids closer together than this distance were excluded.

For the computational "random" model of nucleoid placement, 200 simulations

were run exactly matching the number of nucleoids observed and the lengths of mitochondrial segments for each cell. Nucleoids were placed at a random position in unbranched mitochondrial segments longer than 1 um to match the analysis performed on the actual nucleoid data. The number of spots for each segment was Poisson distributed with average proportional to the length of the segment and the total number of nucleoids. To match the detection of the real spots, any simulated nucleoids closer together than 0.227 um were excluded from the simulated data as well.

Detailed explanation of CDF plots

In Fig. 2.2 B, the CDFs of relative partitioning error of nucleoid segregation, the percent of total nucleoids in each cell half was subtracted from the percent of cytoplasmic volume in that cell half. The absolute value of these errors for both cell halves was summed to make the relative nucleoid segregation error for that cell. To calculate the errors if nucleoids were randomly placed in mitochondria, for each cell a binomial distribution of nucleoids in each cell half was created using the actual number of nucleoids and the actual split in the amount of mitochondria that was segregated to each cell half.

In the inset of Fig. 2.3C, the CDFs of relative error of WT nucleoid segregation were computed as above. The errors for simulated spacings were created by simulating nucleoids on a single 1-d mitochondria according the the distributions in 2.3 C. The number of nucleoids and split of mitochondria was taken from the actual cells.

2.5.7 Yeast strains

All strains were based on the *S. Pombe* strain FWP172 (a gift of F. Winston) except for PHP14 (a gift of T. Fox). Deletions were constructed using standard PCR mediated techniques or PCR techniques based on the *S. Pombe* deletion collection [58].

Plasmid construction

Plasmid pRJ06 was constructed using isothermal assembly [59] of the PCR product of primers RJ-036 (5'-CAG GTG CCT TCG CTT TTC TTT AAG CAA GAG AAT TGT CGA GAT GGC CTC CAC TCG TGT CCT-3') and RJ-037 (5'-GAT GAT GGC CAT GTT ATC CTC CTC GCC CTT GCT CAC CAT GGA AGA GTA GGC GCG CTT CTG-3') on plasmid pYES-mtGFP (gift of B. Westermann [60]) and primers RJ-034 (5'-ATG GTG AGC AAG GGC GAG G-3') and RJP-035 (5'-CTC GAC AAT TCT CTT GCT TAA AGA AAA GCG AAG G-3') on plasmid pDH92 (gift of D. Huh, unpublished). This plasmid was then cut with NotI and transformed into strain DH0 to make RJP005.

Plasmid pRJ26 was made from isothermal assembly of the the PCR product of primers RJ-037 and RJ-094 (5'-AAC ACG GGA TCC CCG GAT CC-3') on plasmid pRJ07 (unpublished) and primers RJ-034 and RJ-117 (5'-AAT TCC TGC AGG ATC CGG GGA TCC CGT GTT ATA TTA CCC TGT TAT CCC TAG CGG ATC TGC-3') on pDH62 (gift of D. Huh, unpublished). This plasmid was cut with MfeI and transformed into yFS284 to make RJP028.

Table 1: Strains used in this study

Name	Genotype	Source
DH0	ade6::ade6+ ura4-D18 leu1-32 ade6-m210 h-	D. Huh
yFS284	ura4-D18 his7-366 cdc25-22 leu1::pFS181(leu1+ adh1:hENT1) pJL218(his7+ adh1:tk) (integrated at random position)	N. Rhind [42]
PHP14	ade6-M216, leu1-32, ptp1-1, ρ^0	T. Fox[14]
DH60	leu1::Padh1-GST-NES-mCherry-leu4+ leu1-32 ura4-D18 h-	D. Huh
RJP005	ura4-D18 leu1-32:pRJ06(leu+:adh1pr:mtmCherry)	This Study
RJP016	ura4-D18 leu1-32:pRJ06(leu+:adh1pr:mtmCherry) mmb1Δ::kanMX4	This Study
RJP025	ura4-D18 leu1-32:pRJ06 (leu+:adh1pr:mtmCherry), pom1Δ::kanMX	This Study
RJP028	ura4-D18 his7-366 cdc25-22 leu1::pFS181(leu1+ adh1:hENT1) pJL218 integrated (his7+ adh1:tk), pRJ26 integrated (pADH1-MLS-mcherry)	This Study
RJP029	ura4-D18 leu+:pRJ06 (leu+:adh1pr:mtmCherry), wee1Δ::KanMX6	This Study
RJP041	ura4-D18 leu1-32:pRJ06 (leu+:adh1pr:mtmCherry), prp43:msfGFP:URA4	This Study
RJP042	ura4-D18 leu1-32:pRJ06 (leu+:adh1pr:mtmCherry), pom1Δ::kanMX, prp43:msfGFP:URA4	This Study
YJ014	prp43-msfGFP-ura4+, leu1::Padh1-GST-NES-mCherry-leu4+ leu1-32 ura4-D18 h-	This Study

Strain construction

PRP43-msfGFP translational fusions (for RJP041 and RJP042) were constructed by colony PCR of YJ014 with primers RJ-157 (5'-AGT GGA TTT TTC ATG CAA GTT GCC-3') and RJ-158 (5'-CAT ATT TTT GGC ATA AAG CTG CAC G-3') ($\sim$ 3.5 kB) and subsequently transformed into RJP005 to make RJP041 and RJP025 to make RJP042 with URA selection.

YJ014 was constructed by transforming the PCR product of msfGFP-ura4 (gift of D. Landgraf) flanked by 300bp homologous sequences for prp43 following the protocol provided by Bahler et al.[61]. The two homologous sequences H1/H2 were amplified from purified DNA from DH0 using a pair of primers YJ014-f1 (5'-AGT GGA TTT TTC ATG CAA GTT GCC-3') and YJ014-r1 (5'-CCA GCA CCA GCA CCT GCT CCA CGT CGA GCG TTC TTT TTT GAT CTA G-3'), and a pair of YJ014-f2 (5'-GTT TAA ACG AGC TCG AAT TCA TCG AAA TCT AAT TTA CTG CTC GGT GAA TTA CAA ATA T-3') and YJ014-r2 (5'-CAT ATT TTT GGC ATA AAG CTG CAC G-3'), respectively. And then, msfGFP-ura4 flanked by the homologous sequences were amplified from pYJ003 and H1/H2 using primers of YJ014-f1 and YJ014-r2.

2.6 Supplemental figures

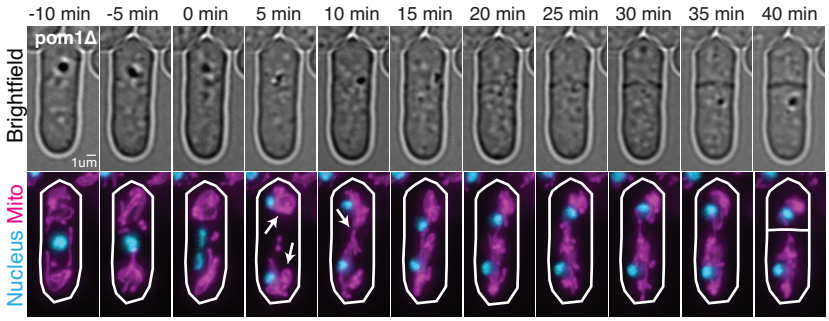

Figure 2.5: Timelapse images of a dividing *pom1*Δ cell (RJP042) in brightfield and with mitochondrial matrix-targted mCherry (magenta) and GFP labeled nuclei (cyan). Arrows at 5 min indicate initial separation of mitochondria at the same time as nuclei and then reformation of a continuous mitochondrial network at 10 min but before division (40 min). Scale bars indicate 1um.

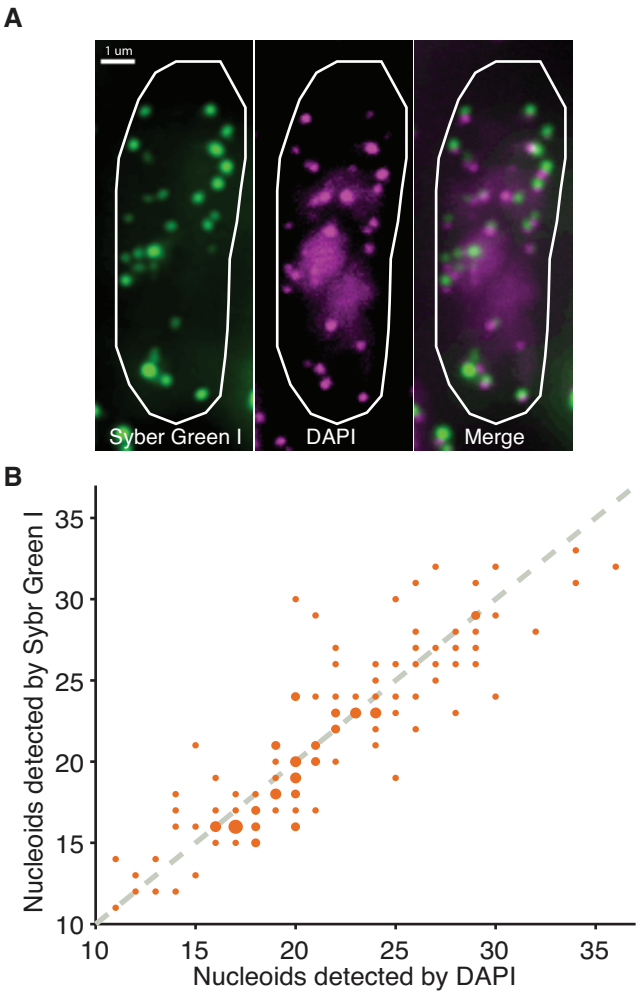

Figure 2.6: Quantification of mitochondrial nucleoids with fluorescent dyes **(A)** Sample images of an *S. pombe* cell stained with Sybr Green I (green) and DAPI (magenta). The outline of the cell is provided for reference. Scale bar is 1um. **(B)** The number of nucleoids detected with SGI or DAPI in each cell is plotted. The area of each point is proportional to the number of samples observed. The dashed gray line indicates equality between the two methods. (n=116 cells (RJP005)).

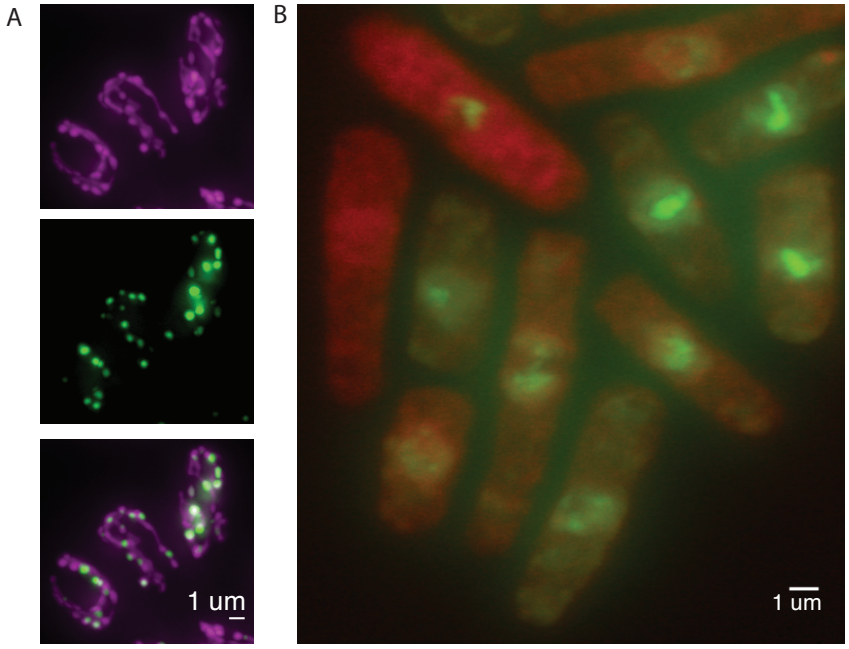

Figure 2.7: Sybr Green I staining is specific to mitochondrial nucleoids. **(A)** RJP005 imaged with Top: Mitochondrially localized mCherry (magenta), Middle: Sybr Green I staining (green). Bottom: overlay. Note that SGI staining was used to determine the number of nucleoids in each cell but the location and morphology of the mitochondria and nucleoids were disturbed. **(B)** ρ^0 cells (PHP14) stained with Sybr Green I (green) and autofluoresence in mCherry channel (red). Only nucleus and no mitochondrial nucleoids are visible.

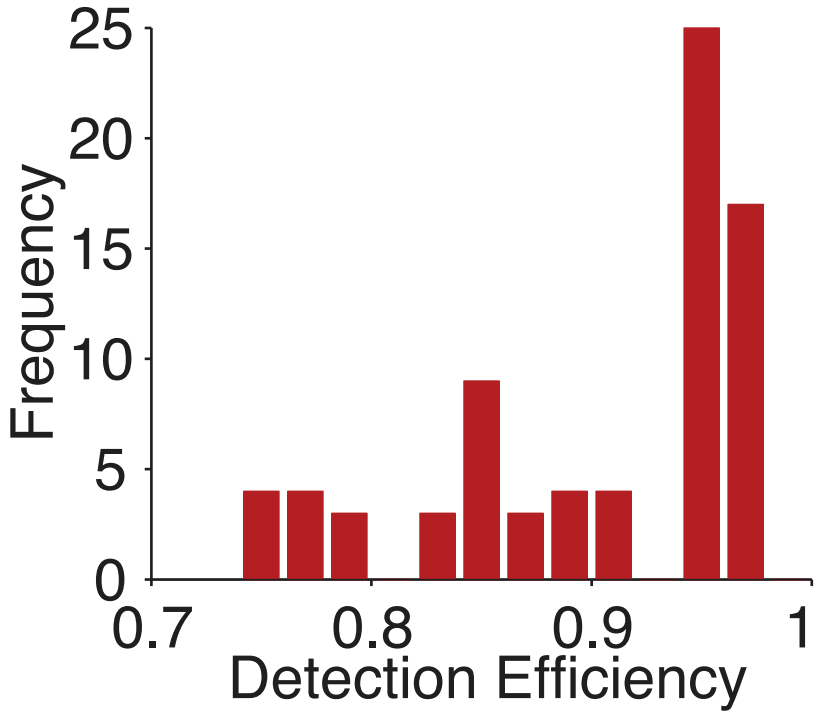

Figure 2.8: Histogram of detection efficiencies calculated by using the variability in detection between SGI and DAPI and assuming a binomial model of nucleoid detection.

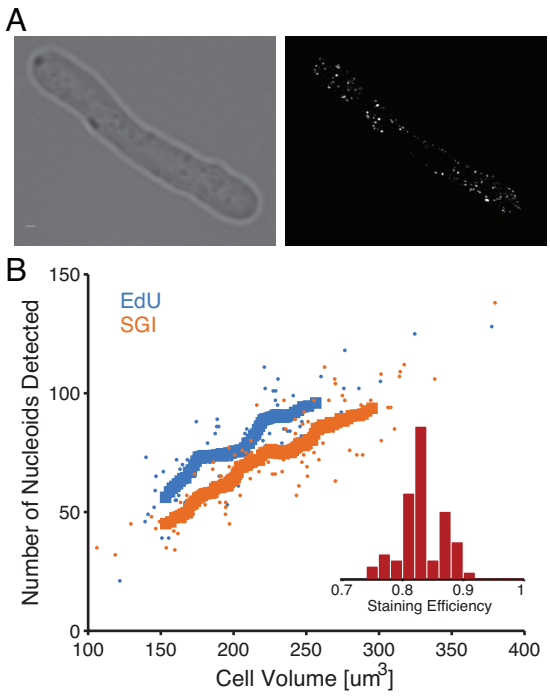

Figure 2.9: EdU staining provides orthogonal method to verify Sybr Green I staining efficiency. **(A)** A yeast strain (ySF284) was arrested in G2, allowed to incorporated EdU, fixed and Alexa-488 azide was ligated to EdU with "click" chemistry (see methods). Left: Bright-field image of cell, Right: Alexa-488 channel visualizing nucleoids. **(B)** The number of nucleoids detected by SGI and EdU plotted against cell volume. Thick lines are running average of 20 cells. Inset: histogram of calculated efficiency of SGI vs EdU. To calculate efficiency, the running average of nucleoids identified by SGI was compared to the running average of nucleoids identified by EdU. For each point in the running average of SGI, the cell volume was noted and a matching cell volume from the running average of EdU cells was chosen for comparison. If there was no exact match in cell volume, a linear interpolation from the two points closest in volume was used. (n=77 cells for EdU and n=108 cells for SGI).

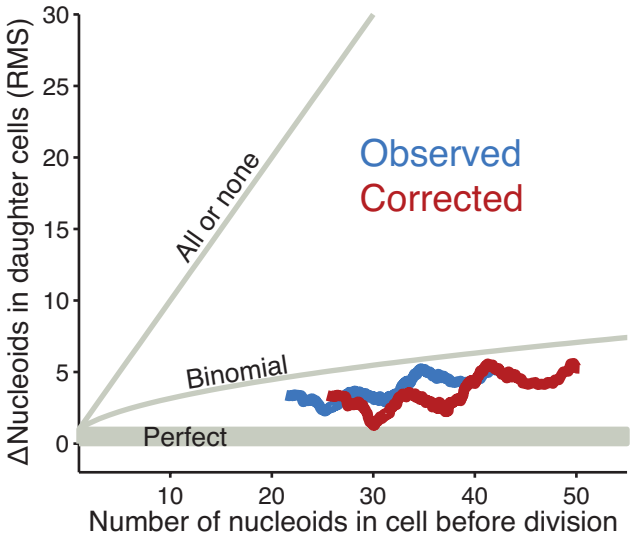

Figure 2.10: Correcting for random undercounting of nucleoids shows that wild-type cells partition even more accurately than observed. A running average (50 points) of the root mean square (RMS) error between wild-type daughter cells is plotted against the total number of nucleoids in the cell before division (blue). The corrected version (red) shifts to the right to correct for 84% undercounting of nucleoids. The line moves down because random undercounting is a binomial process and undoing this, shifts the trend away from the binomial line. See methods and Huh and Paulsson [1] for details.

3

Consequences of cooperativity in yeast polarity and other systems

3.1 Introduction

The spontaneous formation of biological patterns from uniform initial conditions is an enigmatic and essential property of many biological systems. Turing proposed a

general explanation for this phenomenon requiring non-linear positive feedback of morphogens and differential diffusion of activators and inhibitors [17]. Many examples of pattern formation have been attributed to Turing mechanisms [62–64], mostly at the level of plausibility but without explicit proof of the required conditions. In addition many other mechanisms besides diffusion and reaction can spontaneously create patterns [65, 66]. Here we explore whether the non-linearities required by Turing-type mechanisms have trade-offs that might prevent their widespread implementation in biology.

We use the spontaneous bud formation in *Saccharomyces cerevisiae* as a model system for studying pattern formation. Each cell division cycle *Saccharomyces cerevisiae* chooses a single site on the plasma membrane from which to start growing a daughter cell. In wild-type cells new buds use pre-existing landmarks to determine the position of the bud, but even when the ability to use landmarks is disrupted, *S. cerevisiae* still buds at a single site on the cell cortex [67]. The core of the bud site selection network is the GTPase Cdc42p and its effectors, Cdc24p and Bem1p which help Cdc42p cycle between its cytosolic and membrane-bound forms [68] to eventually concentrate in one location on the plasma membrane. Cdc42p then recruits downstream factors needed for budding. Importantly it recruits factors that lead to the establishment of a stable actin network at the site of polarity. This actin network shuttles vesicles containing membrane-bound Cdc42p to and from the membrane and deposits materials required for budding. The net effect of actin is to stabilize the bud site, but the precise mechanism by which actin-based feedback accomplishes this feat is debated [69, 70]. Homologues of Cdc42p and its effectors

are highly conserved and are used for pattern formation in other eukaryotes such as *C. elegans* posterior-anterior specification and mammalian cell migration [71, 72]. Thus understanding pattern formation in the yeast system may also lead to insights to other pattern formation systems.

Most models of the establishment of cell polarity have focused on achieving a single site of polarity [17, 18, 73–75]. Usually, questions center on whether polarity requires a Turing type model and what the minimal requirements are for a gene regulatory network to achieve polarity [64, 71, 76–83]. Here we explore whether other biological criteria such as quickness in polarizing and stability of polarity once it is established additionally constrain the potential networks that can achieve polarity. We mainly focus on a key feature of many polarity models, cooperativity, and how it changes the ability for a polarity system to meet these other biological criteria. We find that the non-linearities required for a Turing-type model lead to slow polarization; the system is severely limited by the finite speed of molecular diffusion and the details of a realistic reaction mechanism. Due to the limited time of the cell cycle, this slowness could be a selective pressure which limits cooperativity in this system.

Although these trade-offs do not rule out a Turing-type reaction mechanisms for bud site selection, they do pose new constraints that may make Turing-type mechanisms less likely or require further mechanisms to make them more robust. Given the current understanding of bud site selection, we propose that the actin-mediated polarity serves as an additional stabilizing mechanism. Distinct from the standard Cdc42p based cooperative feedback mechanism that relies on

cooperativity to make the "bigger spot win" [84], we propose that the actin system achieves a single stable bud-site using a mechanism by which the "older spot wins". We demonstrate the plausibility of this principle with a toy model based on previous observations of actin-based positive feedback. Finally, we extend our analysis to other systems that involve cooperativity, and find the same trade-offs restrict the design of these system. Thus a previously unappreciated trade-off between concentration response and temporal response seems to be a general feature of cooperativity.

3.2 Kinetic trade-offs with cooperativity in polarization

To investigate the implications of cooperative reactions in the formation of polarity, we used a toy model that contains the essential features of the known biology of yeast polarity, derived from a model used by Altschuler et al. [79, 85]. The model consists of an abstracted yeast cell with a cytoplasm enclosed by a membrane (Fig.3.1A). At the start of polarization, the cell contains many copies of a single signaling species, analogous to Cdc42p, isotropically distributed in the cytoplasm. When these molecules are concentrated at a specific location on the plasma membrane, they mark the bud site. Each molecule can spontaneously bind to the membrane at any location and then recruit others from the cytoplasm to that location on the membrane (positive feedback). In addition, molecules in the membrane can dissociate and re-enter the cytoplasm. As shown previously [79], this combination of actions by the signalling molecules results in a dynamic cluster at a single location on the cell membrane (Fig. 3.1B). With non-cooperative rates of

recruitment (linear or $n = 1$ in 3.1A) polarization occurs, but is subject to destabilization from competition from spontaneous binding events at other sites on the membrane [79].

While Altschuler et al. focused on whether a single site of polarity was achievable, we additionally evaluate the time this mechanism takes to polarize. As we are specifically interested in whether there is a trade-off between the degree of cooperativity and the time scale to achieve polarization we simplify the model to eliminate diffusion in the membrane. This assumption makes it easier for stable spots to form, allows a more rigorous definition of what a spot is, and places a lower bound on the formation time. This assumption is reasonable in yeast where diffusion on the membrane is small ($0.036um^2/s$) [86] compared to cytoplasmic diffusion of a typical protein ($\sim 10um^2/s$) [87].

Models testing the effects of cooperativity are often compared by holding all other parameters constant and changing a cooperativity factor [64, 82]. While instructive for understanding cooperativity, these models do not address whether the tested cooperativities are biologically plausible. Cooperativity is reflected in super-linear changes in the reaction rate with changes in concentration of the molecular species, but also cause the maximum reaction rates to rise. To compare the trade-offs due to cooperativity we instead capped the maximum rate of recruitment and assumed that cooperativity is implemented in recruitment. Cooperative recruitment means that molecules on the membrane recruit from the cytoplasm more strongly when surrounded by other molecules on the membrane (i.e. $n > 1$ for the equation in Fig. 3.1A). Although diffusion sets the speed limit for reactions by setting the rate of

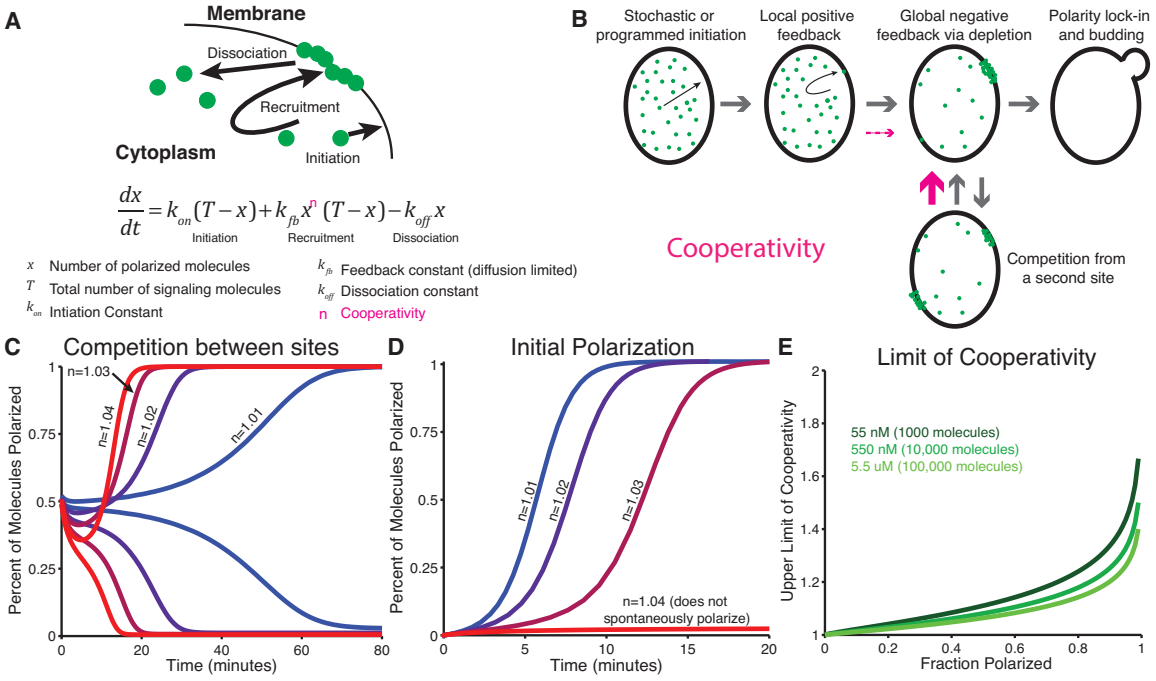

Figure 3.1: A simple model of polarity with cooperativity **(A)** A detailed reaction scheme of recruitment and dissociation of molecules at the plasma membrane along with an equation and parameters mathematically describing the dynamics of the system. **(B)** Schematic of polarization where signaling molecules (green) are either in the cytoplasm or in the membrane. A single spontaneous binding event leads to further recruitment of molecules to that site. While other sites are repressed through depletion of molecules from the cytoplasm, a second site can also be seeded and begin competing with the initial site. Singular polarity can only be established after competition is complete and a single site remains. Cooperativity decreases the time to resolve from two sites to one site, and helps to favor existing sites over new sites, but this benefit comes at the cost of slower initial polarization. **(C)** Competition between two competing sites of polarity is accelerated by cooperativity. Two competing sites of polarity are initiated at nearly the same number of molecules and differing amount of cooperativity (n from (A)) **(D)** The formation of polarity is delayed as cooperativity increases, until polarity is not possible. C and D both assume that $T = 1000$ [79], $k_{off} = 9min^{-1}$ [80] and that the maximum value of recruitment is $3 \cdot 10^6\ s^{-1}M^{-1}$ or $0.01\ min^{-1}molecule^{-1}$ about one hundred times slower than the diffusion limit [88]. **(E)** Given the measured $k_{off} = 9min^{-1}$ the upper limit of possible cooperativity is plotted as a function of the fraction of polarized molecules (see Methods).

molecular collisions, very few enzymes ever achieve this limit [88]. Instead, most collisions are unproductive, occurring at the wrong angle, orientation or speed, meaning that the rate of most reactions is far below the diffusion limit. Thus we took the maximal recruitment rate of our signalling molecules to be at such a physiological rate, about one hundred times lower than the diffusion limit.

In order to achieve a single site of polarity, models must be able to quickly resolve multiple sites to a single site by competition. We therefore first tested how cooperativity effected the timescale to resolve two competing sites. We seeded two sites with nearly equal numbers of signalling molecules and measured how quickly competition occurs using an equation like that in Fig. 3.1A but with two sites (see Methods). As expected the site with more molecules dominates more quickly as cooperativity is increased (Fig. 3.1C). This makes sense because cooperativity is defined by the fact that recruitment grows more quickly than dissociation as the number of molecules increases. This confirms the accepted result that cooperativity is a desirable property for ensuring robust polarity because it quickly ensures the site with more molecules dominates.

A second critical feature of polarity is the amount of time a cell takes to establish a new site of polarity. If this time is too long it would delay the cell cycle resulting in lowered growth rate. To simulate the time cells take to polarize, we determined how long polarization took after a single initial molecule associated with the membrane. In order to appropriately compare the models with the same biophysical limits, we set the maximal rate of recruitment to be the same at the equilibrium polarized concentration for all levels of cooperativity. Thus to increase cooperativity

in recruitment, the rate of recruitment at low concentrations had to be decreased by decreasing the k_{fb} parameter. We found that adding cooperativity to the recruitment reaction then drastically slowed down the recruitment of polarity molecules to the membrane (Fig. 3.1 D). As cooperativity was increased, molecules eventually fell off the membrane faster than they were recruited. Spontaneous polarity was no longer possible at cooperativities below two even though molecules dimerized on the membrane.

In addition, we found that regardless of the number of signalling molecules or the percent of them that polarize, there will be a limit to the amount of cooperativity allowed in this polarization system beyond which spontaneous polarity does not occur (Fig. 3.1 E) . This limit does not depend on the recruitment or dissociation rates but simply that there is a finite upper limit for the recruitment rate and a positive lower limit for the dissociation rate (see Methods). In summary, a toy model for polarity with cooperative recruitment shows that cooperativity helps stabilize an established single site of polarity by speeding competition, but this comes at the cost of slowing initial formation of polarity.

3.3 Mechanistic limitations on cooperativity

The level of cooperativity in a system is often assumed to be the number of components that interact, but this is actually the upper limit of cooperativity. In order to approach this limit, systems have to be far away from saturation or the diffusion limit. We use the toy polarity system detailed above to explore one way that polarity systems are limited in achieving a high level of cooperativity.

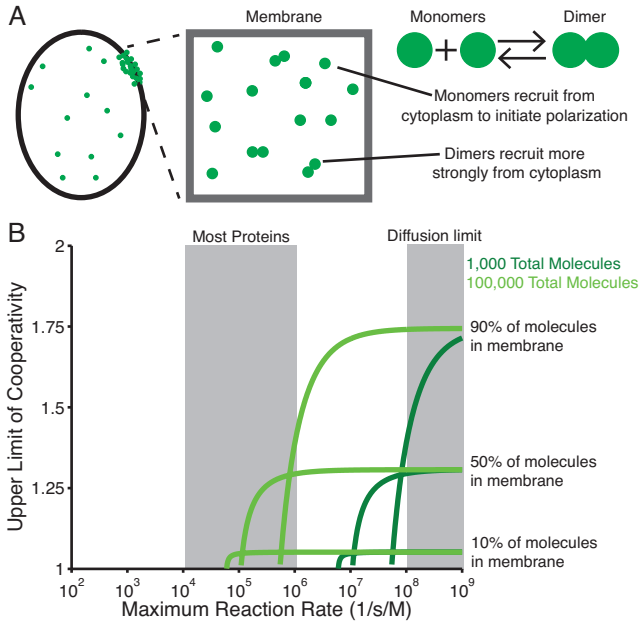

Figure 3.2: Mechanistic limits on cooperativity **(A)** An illustration of a monomer-dimer reversible reaction on the membrane. Signalling molecules can exist only as monomers in the cytoplasm and monomers or dimers in the membrane. This type of reaction allows cooperativity in recruiting molecules from the cytoplasm. **(B)** The average cooperativity allowed is plotted against the maximum recruitment rate for the dimer. Each curve shows a different total number of molecules or percent of molecules polarized at steady state. As before $k_{off} = 9$ min^{-1}.

First we note that the type of cooperative reactions possible in spontaneous pattern formation are inherently different from those in TFs or allosteric ligand binding where a predefined scaffold helps assemble complexes. Cooperativity in polarity must instead occur through interaction between the signaling molecules themselves. One natural way for cells to achieve cooperativity is for signaling molecules to reversibly dimerize with each other on the membrane. Indeed, yeast polarity is believed to involve the formation of a complex between Bem1p and Cdc42p [64]. In our toy model, we allow the single type of signalling molecule to exist as monomers in the cytoplasm and as monomers or dimers on the membrane (Fig. 3.2A). We also assume that monomers and dimers instantly achieve equilibrium in a site of polarity. This approximation allows analytical treatment of the problem to understand its basic features.

This monomer-dimer system can have cooperativity, but is limited by the fact that polarity must be able to form spontaneously. If dimers are able to recruit molecules to the membrane at a higher rate than monomers, the effective recruitment would be super-linear because dimer concentration increases super-linearly with the concentration of signaling molecules on the membrane (Fig. 3.2). However, as the concentration of dimers grow, its rate of increase diminishes – a phenomenon often termed saturation – and the cooperativity starts to diminish. In addition, the recruitment by monomers must be also taken into account: they increase sub-linearly with concentration of signalling molecules, lowering the overall cooperativity. Although this type of dimerization reaction is often abstracted as an idealized non-saturating system with a cooperativity parameter of 2, we find that

the average cooperativity in polarity is far below two in most cases (Fig. 3.2 B) .

We also found that cells could increase the amount of cooperativity possible by changing certain physiological parameters. Increasing the fraction of signalling molecules that polarized at steady-state (Fig. 3.2B) increases the amount of polarity possible. However, given previous estimates for the fraction of signaling molecules in the membrane (39% [89]), the amount of cooperativity is still limited to roughly 1.25 (Fig. 3.2B). Increasing the number of signalling molecules produced by the cell or the rate at which dimers can recruit molecules to the membrane also extend the range of parameters where cooperativity achieves this maximal value. Even taking extremely favorable values for these parameters achieves only limited cooperativity: feedback rates would have to approach the diffusion limit ($10^8 - 10^9$ $s^{-1}M^{-1}$) and 90% of molecules would have to polarize for average cooperativity to be as high as 1.75.

3.4 An alternate mechanism for stabilizing polarity

In yeast budding, cooperativity could ensure that a single bud site dominates, but given the drawbacks of cooperativity and the difficulty in biological implementation, we wanted to test whether properties of the second, actin-mediated feedback loop could enhance polarity.

Previous studies have suggested that the actin feedback loop stabilizes polarity [82] or ensures a single bud site [84], but we show that a previously unappreciated detail of actin-based feedback could also account for the observed phenomena. It has been shown experimentally that actin-based reinforcement of polarity takes cues

from diffusion-mediated spontaneous polarity [84]. As actin patches and filaments form, Cdc42p is delivered to the site of polarity via exocytosis and taken off the membrane via endocytosis. Although the exact mechanism of reinforcement is in question [70, 80, 86, 90, 91], once actin is engaged at a site of polarity, that location turns into a bud and suppresses all other sites not reinforced by actin [84]. A single site of polarity is usually present when actin-based feedback engages as competition between sites leaves a winner very quickly (<1.5 minutes [84]), but in mutant cells where diffusion-mediated competition between sites of polarity is slowed, two sites are sometimes reinforced by actin, resulting in two buds and eventual cell death [84]. Therefore we reasoned that it is not the process of actin-based feedback itself that ensures a single site of polarity. However, Howell et al. also noted that in wild-type cells actin-based feedback does not activate instantaneously after diffusion-mediated polarity; it acts after a narrowly distributed time delay [84]. We therefore investigated whether this type of time delay could cause singularity in polarity.

To understand the properties of a narrowly distributed time-delayed actin feedback loop, we made a conceptual model of how the diffusion-mediated and actin-mediated feedback loops interact. A narrow time delay naturally arises when several sequential reactions are required for a process to begin [92]. This type of time delay makes sense for actin-mediated feedback given the large number of nucleation proteins recruited by Cdc42p to establish actin-based feedback [93]. We compared actin-based feedback that acts after either a single step process or a multi-step process with the same mean delay time; the single step process has a wide (exponential) distribution of delay times while the multi-step process has a

narrower (gamma) distribution of delay times (Fig. 3.3). In both cases, we assumed that the actin mechanism stabilizes an existing spot and destabilizes spots that have no actin feedback (e.g. by increasing the amount of Cdc42p in one spot on the membrane, thereby decreasing the effect concentration of Cdc42p in the cytoplasm). We also assumed that if a site of polarity disappears before actin activation occurred, then actin nucleation proteins also disappeared quickly without establishing an actin network.

We analyzed how the wide and narrow delay distributions for actin-based feedback interacted with the three scenarios that occur as a result of diffusion mediated polarity. In the first case, representing the majority of polarity events observed in [84], diffusion-mediated polarity initiates at a single location which is maintained until actin "locks-in" the bud site. In this case both types of actin feedback loops successfully ensure a single site of polarity (Fig. 3.3). However, when a second site of polarity "invades" the cell before actin is established, a single step actin feedback mechanism might support both sites, leading to two buds. A multi-step actin process only supports the "older" site of polarity because it has been building up the necessary nucleation proteins, which the "newer" site has not had the time to assemble (Fig. 3.3). In the third case, when two sites are seeded at the same time, a single step process might activate before the two sites are finished competing because of the inherently wide distribution of activation times for a single step process (Fig. 3.3 A, histogram). On the other hand, the multi-step process would not activate until all necessary nucleation proteins had been assembled, allowing diffusion-mediated competition to complete before actin

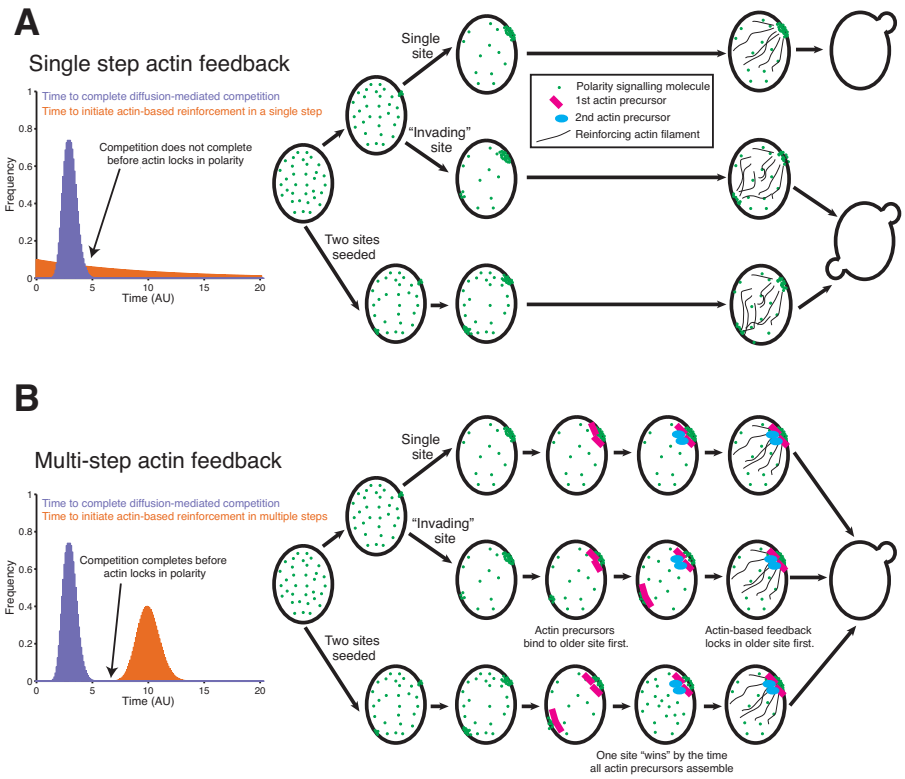

Figure 3.3: A secondary polarity mechanism that works via a time delay can improve polarity (**A**) (Left) Cartoon histogram comparing competition time between sites of polarity (purple) and the activation time for single-step actin mechanism (orange). (Right) Schematic of single-step actin-based feedback showing a best-case scenario when only a single site is seeded, a case where a second site appears after the first and a case where two sites of polarity form simultaneously. In the latter two cases, cells could bud at two locations. (**B**) (Left) Cartoon histogram comparing competition time between sites of polarity (purple) and the tight distribution for multi-step actin activation (orange). (Right) Schematic of multi-step actin-based feedback showing the three cases from (A). In all cases cells will bud at only one site.

locked-in the "winning" site of polarity (Fig. 3.3B, histogram).

3.5 Cooperativity and kinetics in other systems

Cooperativity exists in many other biological systems, and given the trade-offs we found in a model of yeast polarity, we wanted to determine if other systems are also limited by similar trade-offs. Specifically, we analyzed how multi-site transcription factor (TF) binding and, multi-site ligand binding might exhibit a trade-off between sensitivity to concentration and temporal response to concentration changes, two desirable properties for these systems (Figure 3.4A). In our toy model, transcription of mRNA for a single gene is controlled by TFs binding to the promoter; three identical TF binding sites must be occupied for a gene to be transcribed. In such a system transcriptional response can depend sensitively on TF concentration depending on the individual binding and unbinding constants, specifically that TFs bind more easily when other TFs are also bound (Figure 3.4 B). This scheme is used in genes that need an all-or-none response based a concentration threshold of a TF. In the other toy system, an oligomeric enzyme binds ligands more tightly when other sites are already occupied by ligands. This is the type of mechanism behind hemoglobin's ability to carry and release oxygen [94]. Although both systems are biologically distinct, the underlying reaction schemes are the same, allowing us to analyze both types of systems with the identical mathematical framework. We therefore refer to the TF case below, but the conclusions hold for both systems.

We simulated the TF system with different levels of cooperativity by changing binding and unbinding constants and found that its response time to changes in

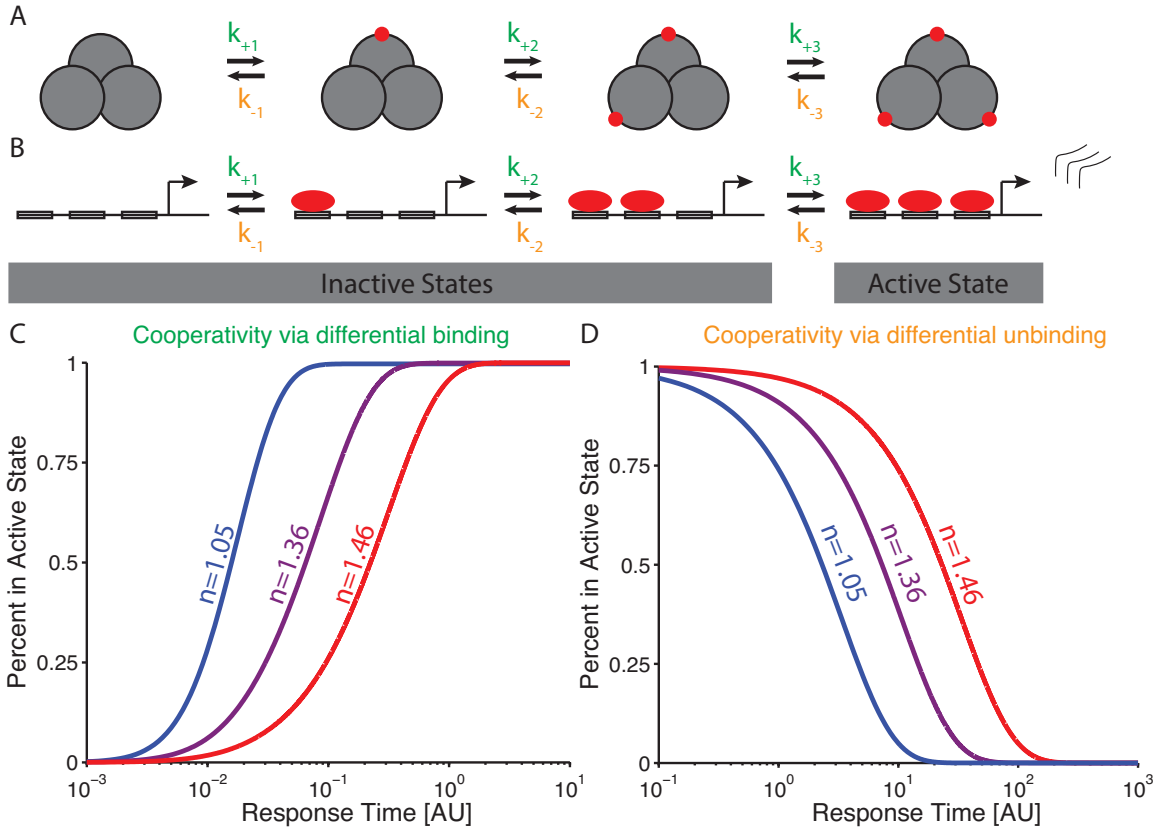

Figure 3.4: Sensitive response to concentration causes delayed response to changes in concentration in other biological systems. (**A**) Schematic of a toy multimeric enzyme system where the binding and unbinding of identical ligands to identical sites is controlled by the rate constants k_i and the enzyme is only active when all sites are bound. (**B**) Schematic of a toy TF-promoter system where the binding and unbinding of identical TFs to identical sites is controlled by the rate constants k_i and gene is only transcribed when all TF binding site are occupied (**C**) A simulation of the response of the TF-promoter system in (B) to a step increase to the TF concentration at t=0. Holding all unbinding rates constant and changing binding coefficients of TFs yield different cooperativities (n) and different delays in response as shown. $k_{-1} = k_{-2} = k_{-3} = k_{+2} = 1$ and $k_{+3} = 10^1, 10^{1.5}, 10^2$ for increasing cooperativity. k_{+1} was set by ensuring that the promoter is in the active state 50% of the time when the TF concentration is 1. (**D**) Similar to (C) except that the response to a step decrease to TF concentration is shown and unbinding coefficients were changed to modify cooperativity while binding coefficients were held constant. $k_{+1} = k_{+2} = k_{+3} = k_{-2} = 1$ and $k_{+3} = 10^{-1}, 10^{-1.5}, 10^{-2}$ for increasing cooperativity. k_{-1} was set by ensuring that the promoter is in the active state 50% of the time when the TF concentration is 1.

concentration slowed with increasing cooperativity (Figure 3.4B and C). Either binding or dissociation rates can be changed to achieve cooperativity: if binding rates are changed to achieve cooperativity, responses to increases in concentration are delayed; if dissociation rates are changed to achieve cooperativity, responses to decreases in concentration are delayed. These results can be rationalized in the following way: a cooperative multi-step reaction like the one we are analyzing achieves cooperativity by making the first step slow and the subsequent steps fast. The slow steps control the cooperativity but also include the rate limiting step for the reaction. The slower the first steps, the more cooperative the system but the slower that changes in the input will be reflected in the output. This finding has implications in many types of control systems as they would be sensitive to the concentration of inputs but would act with a delay. The fact that these types of circuits introduce delays could also undermine their usefulness, as delays in feedback systems often fail to provide robust control [37].

Although this trade-off appears in the simple system we analyze, promoters and oligomeric enzymes, like the polarity system, often include additional methods of regulation. Both of these systems can use energy to speed the transition into the fully bound or empty states to overcome the trade-offs inherent in employing cooperativity. For example phosphorylation of transcription factors or allosteric binding of signaling molecules to a complex can rapidly shift the equilibrium of states, allowing fast changes in the state of the system despite the trade-offs we showed above.

3.6 Discussion

In this study, we have shown some previously unappreciated consequences of cooperativity in spontaneous polarity in a simple budding yeast system. The limits and trade-offs we have identified do not explicitly rule out cooperativity, especially since not all polarity proteins have been identified and very few rates have been explicitly measured. However, the results above show that if the level of cooperativity is high, many of the current assumptions about polarity would have to be revised.

Our time-delayed actin model is speculative but does make testable predictions and would be a new way that cells could maintain singularity in polarity while still polarizing quickly. Howell et al. showed some evidence that the time delay between the two feedback loops is narrowly distributed, and that competition between sites resolves more quickly than this delay time [84], but more measurement of delay times and competition times is warranted.

The fact that other cooperative systems have delays in temporal response is a previously unappreciated phenomena. The trade-offs detailed above lend themselves to tests using variations in promoter sequences of TF binding sites. For cooperative ligand binding, mutations in oligomeric enzymes may also allow tuning of the cooperativity of these systems and *in vitro* assays could be used to judge cooperativity and time response to changees in concentration.

3.7 Methods

3.7.1 Limits on Cooperativity in Polarity

Competition between two sites of polarity is described by the set of linked differential equations of the time evolution of the number of proteins at each site (x_1 and x_2):

$$\frac{dx_1}{dt} = k_{fb}{x_1}^n(T - x_1 - x_2) - k_{off}x_1 + k_{on}(T - x_1 - x_2)$$

$$\frac{dx_2}{dt} = k_{fb}{x_2}^n(T - x_1 - x_2) - k_{off}x_2 + k_{on}(T - x_1 - x_2)$$

with parameters, $T = 1000$, total number of molecules in system; $k_{off} = 9$ $minutes^{-1}$, $k_{on} = 10^{-3}$ $molecules^{-1}$ per site. For the feedback coefficient, the maximum feedback (k_{fbMAX}) was set to be 0.01 molecules recruited from cytoplasm per minute per molecule in cytoplasm per molecule in membrane (or $3 * 10^6$ $molecules^{-1}mol^{-1}seconds^{-1}$). To set k_{fb} we assume that the maximum recruitment rate is acheived when there are x_{ss} molecuels on the membrane. This would require that $k_{fb} = \frac{k_{fbMAX}}{x_{ss}^{n-1}}$ to ensure that x_{ss} molecules eventually polarized to a single site. x_{ss} was set to 100 molecules in all simulations. These parameters were based on parameters used by Altschuler and co-workers [79, 83]. For competition the Matlab O.D.E. solver *ode45* was employed with initial conditions $x_1 = 51$ and $x_2 = 49$. For feedback rate, the diffusion limit is $10^8 - 10^9 mol^{-1}s^{-1}$ [88] and the volume of yeast was assumed to be 30 femtoliters [87].

Time to spontaneous polarization was calculated with the same parameters, and a single site of polarity was simulated with the dynamics

$$\frac{dx}{dt} = k_{fb}x^n(T-x) - k_{off}x + k_{on}(T-x)$$

and initial condition $x = 0$.

To derive the upper limit of cooperativity in this system we first assume that k_{on} is small, specifically that $k_{on}(T-x) \ll k_{off}x$ and $k_{on}(T-x) \ll k_{fb}x^n(T-x)$ for all relevant x values. Most parameters that allow polarization, including the ones used above, also require this condition [79, 85].

Cooperativity can be either implemented in the rate of recruitment or the rate of dissociation. In other words, a site of polarity can recruit super-linearly as a function of the number of molecules in the site or molecules can dissociate from the site sub-linearly as a function of the number of molecules. We consider a cooperative system with constant non-saturating cooperativity in both feedback and dissociation, with cooperativity coefficients n and m respectively. An average higher than 1 for these values indicates cooperative behavior. We thus have the following differential equation:

$$\frac{dx}{dt} = k_{fb}x^n(T-x) - k_{off}x^{2-m}$$

Note that for $(n+m)/2 < 0$, the system does not polarize, for $0 < (n+m)/2 < 1$, the system is anti-cooperative and favors– in the case of many possible bud sites– a uniform distribution of polarity molecules to all sites, for $(n+m)/2 = 1$ displays

neutrally stable bud sites and for $(n+m)/2 > 1$ displays cooperativity that ensures a single site of polarity.

The following two assumptions create a limit for cooperativity: for the recruitment rate we assume there is a maximum recruitment rate per molecule, $max(k_{fb}) = k_{fbMAX}$ such that $k_{fb}x^n \leq k_{fbMAX}x$. This is a natural assumption since rates of recruitment are limited by the diffusion. This limit will eventually set an upper limit on n. For the dissociation rate, we assume that there is a lower limit for the dissociation rate per molecule $min(k_{off}) = k_{offMIN}$ such that $k_{off}x^{2-m} \geq k_{offMIN}x$. This ensures that the lifetime of molecules on the membrane cannot become arbitrarily large. This limit is applicable for two reasons: 1) short times spent on the membrane ensure finite regions of polarity through limited time for diffusion on the membrane [79], and 2) it ensures fast competition times between two sites of polarity. This limit will eventually set an upper limit on m.

To maximize cooperativity,

$$k_{fb}x_{ss}^n = k_{fbMAX}x_{ss} \tag{3.1}$$

and

$$k_{off}x_{ss}^{2-m} = k_{offMIN}x_{ss} \tag{3.2}$$

Intuitively, this ensures that cooperativity is not wasted when $x > x_{ss}$ since that regime will not be explored during spontaneous polarization from $x = 0$.

In addition, we require that x_{ss} molecules polarize at steady state so

$$k_{fb}x_{ss}^{n}(T - x_{ss}) = k_{off}x_{ss}^{2-m} \tag{3.3}$$

Then substituting in equations 3.1 and 3.2 to 3.3 gives

$$\frac{k_{fbMAX}}{k_{offMIN}} = \frac{1}{T - x_{ss}} \tag{3.4}$$

Further, taking the ratio of 3.1 and 3.2 gives

$$\frac{k_{fb}}{k_{off}}x_{ss}^{n+m-2} = \frac{k_{fbMAX}}{k_{offMIN}} \tag{3.5}$$

and combining 3.4 and 3.5 gives

$$\frac{k_{fb}}{k_{off}} = \frac{x_{ss}^{2-n-m}}{T - x_{ss}} \tag{3.6}$$

For spontaneous polarity to occur, the site of polarity must be able to recruit new molecules when it only has c molecules that spontaneously adhere to the membrane (where $c < x_{ss}$ and usually close to 0), so $\frac{dx}{dt} > 0$ for $x = c$ or

$$k_{fb}c^{n}(T - c) > k_{off}c^{2-m} \tag{3.7}$$

Substituting in 3.6 and rearranging gives

$$\left(\frac{x_{ss}}{c}\right)^{2-n-m} > \frac{T - x_{ss}}{T - c} \tag{3.8}$$

Isolating n and m then gives the upper limit for their sum:

$$n + m < 2 + \frac{\log \frac{T-c}{T-x_{ss}}}{\log \frac{x_{ss}}{c}} \tag{3.9}$$

In Figure 3.1E of the main text, we assume $c = 1$ to allow polarity to occur from a single spontaneous binding event.

3.7.2 Cooperativity in Monomer-Dimer Systems

To understand a mechanistic realization of cooperativity in polarity, we used monomer-dimer reactions on the membrane. Instead of constant cooperativity, the recruitment of new molecules is a function $f(x)$ and the dissociation is a constant yielding:

$$\frac{dx}{dt} = f(x)(T - x) - k_{off}x + k_{on}(T - x) \tag{3.10}$$

We assume that the following reaction occurs on the membrane

$$m + m \underset{m^2 k_{-1}}{\overset{dk_{+1}}{\rightleftharpoons}} d$$

where m are monomers that are recruited to the membrane and d are dimers that are reversibly formed from dimers. Therefore $m + 2d = x$, and we assume that monomers and dimers equilibrate quickly compared to the fluctuations in concentration of x on the membrane.

Assuming this conservation and instant equilibration between monomers adn dimers we then have

$$m^2 = k_m \frac{(x-m)}{2}$$

where $k_m = \frac{k_{+1}}{k_{-1}}$ and solving for $m(x)$ and $d(x)$ gives

$$m(x) = \frac{1}{4}(-k_m + \sqrt{k_m^2 + 8xk_m})$$

and

$$d(x) = \frac{1}{2}\left(x - \frac{1}{4}(-k_m + \sqrt{k_m^2 + 8xk_m})\right)$$

If monomers recruit from the cytoplasm with rate k_{fbm} and dimers with rate k_{fbd}, then

$$f(x) = k_{fbm}m(x) + k_{fbd}d(x) \tag{3.11}$$

For spontaneous polarity to occur, a site of polarity must be able to recruit new molecules when it only has 1 molecule that spontaneously associates with the membrane: $\frac{dx}{dt} \geq 0$ for $x = 1$. (In order to maximize the cooperativity of the system, the monomers should recruit as minimually as possible, so we take the equality case.) In this case, the number of monomers is 1 and the number of dimers is 0 so $f(x) = k_{fbm}$ and we further assume that $k_{on}(T-x) \ll k_{off}x$ and $k_{on}(T-x) \ll f(x)(T-x)$ similarly as before so Eqn. 3.10 becomes

$$k_{fbm} = \frac{k_{off}}{(T-1)} \tag{3.12}$$

Then in order for x_{ss} molecules to polarize to the membrane at steady state,

$$f(x)(T - x_{ss}) = k_{off} x_{ss} \tag{3.13}$$

Then combining 3.13, 3.12 and 3.11, we can solve for k_m as

$$k_m = \frac{(T-1)^2 x_{ss}(-2k_{off} + k_{fbd}T - k_{fbd}x_{ss})^2}{k_{off}(-k_{fbd} - 2k_{off} + k_{fbd}T)(T - x_{ss})(x_{ss} - 1)}$$

To find a numerical value for k_m one substitutes in for T, k_{fbd}, x_{ss}, and assumes $k_{off} = 9min^{-1}$ as previously done [79]. We then substitute k_m and the other constants back in to $f(x)$. We then define the average cooperativity of $f(x)$ as the average sensitivity of $f(x)$ to changes in x or

$$\frac{1}{x_{ss}} \sum_{x=1}^{x_{ss}} \frac{df(x)}{dx} \frac{x}{f(x)}$$

which is plotted numerically in Figure /reffig:MonomerDimerB. Unstable solutions are not plotted. Conversion of units for k_{fbd} was done as mentioned in the previous subsection.

3.7.3 Multi-site Cooperativity

In order to simulate transcription factor or ligand binding on multiple sites we consider the following reaction scheme:

$$S_0 \underset{S_1 k_{-1}}{\overset{3xS_0 k_{+1}}{\rightleftharpoons}} S_1 \underset{2S_2 k_{-2}}{\overset{2xS_1 k_{+2}}{\rightleftharpoons}} S_2 \underset{3S_3 k_{-3}}{\overset{xS_3 k_{+3}}{\rightleftharpoons}} S_3$$

where S_i indicates the promoter with i binding sites occupied, x indicates the concentration of TF and k_i's are reaction rates. The only active state for transcription is S_3. Note that statistical factors that indicate the number of ways certain reactions can be completed are also included.

Using detailed balance, we solve for the fraction of the time the system occupies S_3 at equilibrium.

$$S_3(x) = \frac{k_{+1}k_{+2}k_{+3}x^3}{k_{-1}k_{-2}k_{-3} + 3k_{-2}k_{-3}k_{+1}x + 3k_{-3}k_{+1}k_{+2}x^2 + k_{+1}k_{+2}k_{+3}x^3}$$

When allowing the on rates (k_{+i}) to be adjusted to maximize cooperativity, we fix the off rates (k_{-i}) as $k_{-1} = k_{-2} = k_{-3} = k_{+2} = 1$. In addition, we require that when $x = 1$, $S_3 = \frac{1}{2}$ and use this condition to eliminate k_{+1}. We then compute the cooperativity (sensitivity) of the reaction at reaction midpoint, $x = 1$, yielding

$$\frac{3}{2} - \frac{3}{k_{+3}} - \frac{3}{2k_{+2}k_{+3}}$$

When allowing the off rates (k_{-i}) to be adjusted to maximize cooperativity, we fix the on rates (k_{+i}) as $k_{+1} = k_{+2} = k_{+3} = k_{-2} = 1$. In addition, we require that when $x = 1$, $S_3 = \frac{1}{2}$ and use this condition to eliminate k_{-1}. We then compute the cooperativity (sensitivity) of the reaction at reaction midpoint, $x = 1$, yielding

$$\frac{3}{2} - 3k_{-3} - \frac{3}{2}k_{-2}k_{-3}$$

In either case, the system's time response is modeled with the usual differential equations based on the reaction scheme above:

$$\frac{dS_0}{dt} = S_1 k_{-1} - 3xS_0 k_{+1}$$
$$\frac{dS_1}{dt} = 2S_2 k_{-2} - S_1 k_{-1} + 3xS_0 k_{+1} - 2S_1 x k_{+2}$$
$$\frac{dS_2}{dt} = 3S_3 k_{-3} - 2S_2 k_{-2} + 2S_1 x k_{+2} - S_2 x k_{+3}$$
$$\frac{dS_3}{dt} = -3S_3 k_{-3} + S_2 x k_{+3}$$

4

Data-driven models of RNA polymerase dynamics

4.1 Introduction

Transcriptional regulation often occurs through the recruitment of RNA polymerase II (RNAP) to promoters and the release of the polymerase from the promoter region

to productive elongation. However, recent evidence has shown that RNA polymerase is also regulated during transcription elongation via pausing, stalling and backtracking [31, 95–97]. Although there has been some theoretical work on sequence-dependent pausing [28], experimental evidence that RNAP dynamics are regulated directly by sequence is lacking. Others have studied RNAP using single molecule methods and have been able to bring great insight to its enzyme kinetics [22, 98–101], but the small number of sequence templates and *in vitro* nature of the experiments makes it difficult to discern general principles of sequence-related polymerase regulation. Standard RNA sequencing experiments do not give insight to nascent RNA, as nascent RNA constitutes the minority of sequencing reads; most reads are from completed transcripts.

A new method has recently been created to monitor the position of actively trascribing RNA polymerases genome-wide in the budding yeast *Saccharomyces cerevisiae* [102]. This technique, dubbed NET-seq (Nascent Elongating Transcript sequencing), works by immunoprecipitation of RNAP that is still attached to nascent RNA. The 3' end of the nascent RNA is then isolated, sequenced and aligned to the yeast genome. This process reports single-base resolution and strand specific density of RNAP at a population level. These data have been used to understand promoter directionality, RNA splicing and RNA pausing due to nucleosome roadblocks [102]. Surprisingly polymerases stall and backtrack ubiquitously throughout the genome [102], making it likely that sequence could play a role in polymerase dynamics. The single base resolution of the data allowed us to build a biophysical model for the sequence-mediated causes of RNA polymerase

pausing and backtracking. Using the backtracking model, we synonymously recoded sections of two genes in yeast to have different RNAP dynamics in these sections and find that RNAP densities in these regions were indeed changed.

4.2 NET-seq captures information about RNAP pausing and backtracking

Transcription is a complex process that transfers information from DNA into mRNA, but at the molecular level, studies have shown that RNAP does not incorporate new bases at the 3' end of the nascent RNA by simple regular movement along the DNA in one direction [22, 103]. Instead, it often pauses and moves upstream from the 3' end of the growing RNA, doing a random walk along the RNA:DNA hybrid, sometimes extruding the 3' end of the nascent RNA out of the active site of RNAP [104, 105]. In these cases, the enzyme TFIIS (encoded by *DST1* in yeast) stimulates RNAP to cleave the 3' end of the nascent transcript and then restart transcription from this upstream point.

In their original study, Churchman and Weissman showed that both the initial site of polymerase stalling and the subsequent site of nascent transcript cleavage are discernible via NET-seq [102] (Figure 4.1 A). When NET-seq is performed on wild-type cells, the predominant state of the polymerase is in the backtracked post-cleaved state. The large fraction of polymerases that are paused indicate that RNAP spends a large fraction of time in this state, making it important to understand backtracking. When *DST1* is deleted, the 3' end of nascent transcripts is largely left intact even after the polymerase backtracks, and in this case NET-seq

reveals the initial site where RNAP begins backtracking (Figure 4.1A). By performing NET-seq on both WT and *dst1*Δ strains, the positions of both the initial site of backtracking and the site at which backtracking stops is recorded for the entire genome (Figure 4.1 B). We first investigated the initial sites of pausing (*dst1*Δ NET-seq) and then the sites at which polymerase stops backtracking (WT NET-seq).

4.3 Metric for evaluating models and reproducibility of data

In order to build a metric by which to evaluate different models of RNAP dynamics, we first annotated each position in the genome as either having a high number of reads mapped (paused) or a low number of reads mapped (not paused) for each dataset. Even though the distribution of polymerase density was not bimodal, this procedure allowed us easy evaluation of models and comparison with previous work. An arbitrary cutoff of three standard deviations above the local background was chosen as in [102].

Then in order to quantify how well models predicted RNAP density we used the area under the receiver operating characteristic curve (AUC). The receiver operating characteristic curve plots the false positive rate against the false negative rate for a scalar real-valued classifier that predicts a binary output (high or low at a certain position in the genome). If the classifier cannot discriminate between the binary outputs, the rate of false positives equals the rate of false negatives, resulting in a straight diagonal line in the receiver operating characteristic curve and an AUC

of 0.5. In other words, the classifier has no information about the binary output. On the other hand if the classifier can exactly determine whether RNAP will pause at a certain base pair in the genome – i.e. there exists a specific cutoff value for the classifier that perfectly divides the paused and not paused outcomes – the AUC is 1.0. Another intuitive interpretation of the AUC is the probability that a randomly drawn true positive example has a higher valued classifier than a randomly drawn true negative example. Classifiers that partially explain the output receive an intermediate AUC. In this way, it is possible to compare radically different models of polymerase pausing with a single metric based on their power to predict polymerase pausing.

Before we tested biophysical models of RNAP pausing, we tested whether variation from biological or technical sources in the NET-seq technique might overshadow any predictive model of RNAP pausing. We reasoned that if NET-seq data from biological replicate samples showed polymerases pausing in the same locations as the first sample, then the information about the causes of pausing would also be present in the data. We therefore obtained replicate NET-seq datasets for both *dst1*Δ and wild-type strains from the Churchman lab, gathered independently from the initial published data. In order to judge how closely the two samples matched, we used the density of polymerases in the new dataset as classifiers for previously determined pauses in the original sample. We found that the AUC was 0.70 between the *dst1*Δ datasets and 0.84 between the wild-type datasets. Although this means that there is significant noise in NET-seq data, there is also enough consistency that predictive models could give insight into RNAP

dynamics. It also means that a perfect prediction model would have an AUC of 0.70 for *dst1*Δ or 0.84 for the wild-type, not 1.0. Any model that gave a higher AUC than the biological replicates would likely be over-fitting the data.

4.4 Nucleosomes are not a dominant cause of RNA polymerase pausing

To set a benchmark for any biophysical sequence-dependent causes for RNAP pausing, we first wanted to understand how much polymerase pausing could be explained by the hypothesized nucleosome roadblock mechanism proposed by Churchman and Weissman [102]. In this model, DNA-wrapped nucleosomes cause elongating RNAP to pause while the DNA is being unwrapped from the nucleosome. The evidence presented in their work showed a clear signal that nucleosomes were involved in pausing, but the effect could have been minor and still detected due to the large sample size.

To evaluate the magnitude of the effect of nucleosomes on initial RNAP pausing, we used the same map of nucleosome positions as in the original study [106] and aligned them against the RNAP pause locations in the *dst1*Δ strain. We observed that RNAP pausing frequency oscillates with a $\sim$150 base pair period relative to the center of mapped nucleosomes, indicating that nucleosome position does indeed cause some portion of RNAP pausing and that nucleosomes are periodically spaced (Fig. 4.2 B). However, the magnitude of the variation in pausing probability is quite low, less than one percent. We used the distance to the center of the nearest nucleosome as the classifier to predict pausing and found that the AUC of the

nucleosome position classifier is 0.51 (4.2 B), indicating that the mapped nucleosomes cause a small percentage of pausing in yeast. This means that there are further reasons for polymerase pausing yet to be uncovered.

4.5 Sequence dependent energetics cause initial RNA polymerase pausing

We asked whether the simple energetics of base pairing could be the cause of initial RNAP pausing as observed in the NET-seq data of *dst1*Δ strains. The transcription bubble breaks DNA:DNA base pairs ahead of the polymerase and forms RNA:DNA bonds inside the polymerase. On average RNA has slightly higher affinity for DNA that DNA does for DNA, but the precise value varies according to sequence [107, 108]. We hypothesized that positions where RNAP paused simply had higher energetic stability and would be favored thermodynamically. We used the sequence dependent value of RNA:DNA and DNA:DNA base pairing [107, 108] at each position along the genome and averaged this calculation for locations where RNAP paused. We found that pause locations were preceded by regions where the RNA:DNA base-pairing was stronger than the DNA:DNA base-pairing, and they were followed by regions where the DNA:DNA base-pairing was stronger than the RNA:DNA base-pairing (Figure 4.2 C). When we overlayed this computation against the RNAP transcriptional bubble, we found a striking correspondence (Figure 4.2 C). The region behind the pause location corresponded to the RNA:DNA hybrid that would occur inside RNAP. The region where DNA:DNA base pairing was stronger was in the region where this base pair would have just

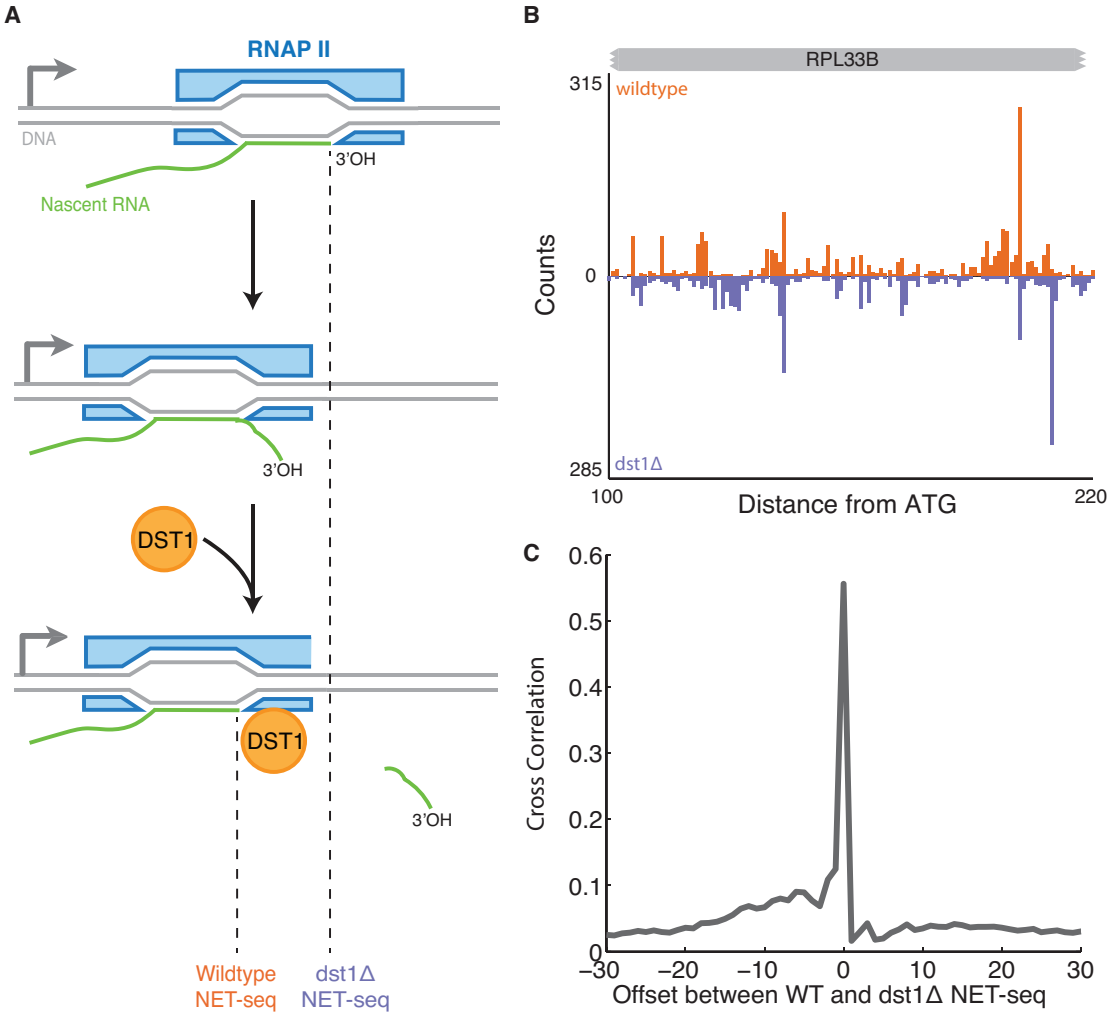

Figure 4.1: NET-seq captures genomic positions of initial RNAP arrest with nucleotide resolution in *dst1*Δ strains and positions after subsequent backtracking and 3' RNA cleavage in WT strains. (**A**)Schematic of RNAP initial pausing and subsequent backtracking; after initial arrest and backtracking, DST1p stimulates cleavage of the 3' end of the nascent RNA. Figure based on [102]. (**B**) WT and *dst1*Δ NET-seq data at the RPL33B gene showing that WT pause locations occur at positions behind *dst1*Δ pauses. (**C**) Cross correlation between WT and *dst1*Δ NET-seq data for highly-transcribed genes (average >2 reads per bp) in coding regions, normalized for level of transcription.

been broken. Thus for RNAP to move forward at this location would require replacing strong bonds with weaker ones, an energetically unfavorable operation. We note that sequences further upstream where the template DNA strain reanneals to the coding strand were not taken into account as little sequence information was shown to be at this location.

We used the biophysical model above to create a predictive metric by adding the energetic advantage of RNA:DNA over DNA:DNA in the eight base pairs upstream of a pause site and the energetic advantage of DNA:DNA over RNA:DNA in the nine positions downstream of the pause site. We then computed this metric for each position in genic regions and found that it correlated with the pause frequency of the polymerase in the *dst1*Δ strain (Figure 4.2 D) . Interestingly pausing seems to rapidly increase after the energy metric is higher than a few kcal/mol, roughly the thermal energy inside a cell. The range of pause frequencies spanned by this metric was 10%, about ten times the range of the nucleosome classifier, indicating that it is more predictive of RNAP pausing. The AUC for the energy metric was 0.64 and therefore does not explain all pausing, but does approach the AUC of the biological replicate (0.70).

We do not understand if there is a functional reason for RNAP to pause at certain energetically determined sequences, but we hypothesize that pausing at these sequences is a trade-off for the stringent error detection of RNA base incorporation. When RNAP misincorporates an RNA base, it detects this error energetically as the mispaired RNA:DNA hybrid will contribute less free energy to the forward movement of RNAP than breaking the next (correctly paired) DNA:DNA base pair.

The polymerase would then be likely to pause and begin backtracking to eventually excise the mis-incorporation with the aid of TFIIS [29, 109–111]. We hypothesize that *any* energetically unfavorable forward movement would also trigger pausing, and sequences of DNA could mimic the energetic consequences of mispaired bases, thus triggering ubiquitous pausing as seen in the *dst1*Δ NET-seq data.

4.6 Causes of polymerase pausing after backtracking

We then asked why RNAP pauses after the 3' end of the sequence has been clipped. These backtracked positions are reflected in the wild-type NET-seq data and show that RNAP has a clear preference to be clipped at certain sequences [102] (Fig 4.3 A), but no explanation exists for why those specific sequences would be associated with high polymerase density.

Using the reported structure of paused RNAP [104] (Fig. 4.3 B), we reasoned that the +1 position influenced polymerase pausing because of RNA polymerase stacking interaction with the nascent RNA. While the polymerase is backtracking, the 3' end of the nascent RNA is thought to interact with the "gating" tyrosine of RNAP (Y769 on *Rpb2p*) as it is extruded out of the pore. We used calculations of the interaction energy between each of the four RNA bases stacked against tyrosine to estimate the energy of interaction of each one [112, 113]. Although the energy calculations were done assuming gas phase isolated molecules, instead of solvated structures, adenine was shown to stack most energetically favorably with tyrosine, perhaps explaining why it occurs toward the 3' end of the RNA as seen in the web-logo at position +1. (Fig. 4.3 A).

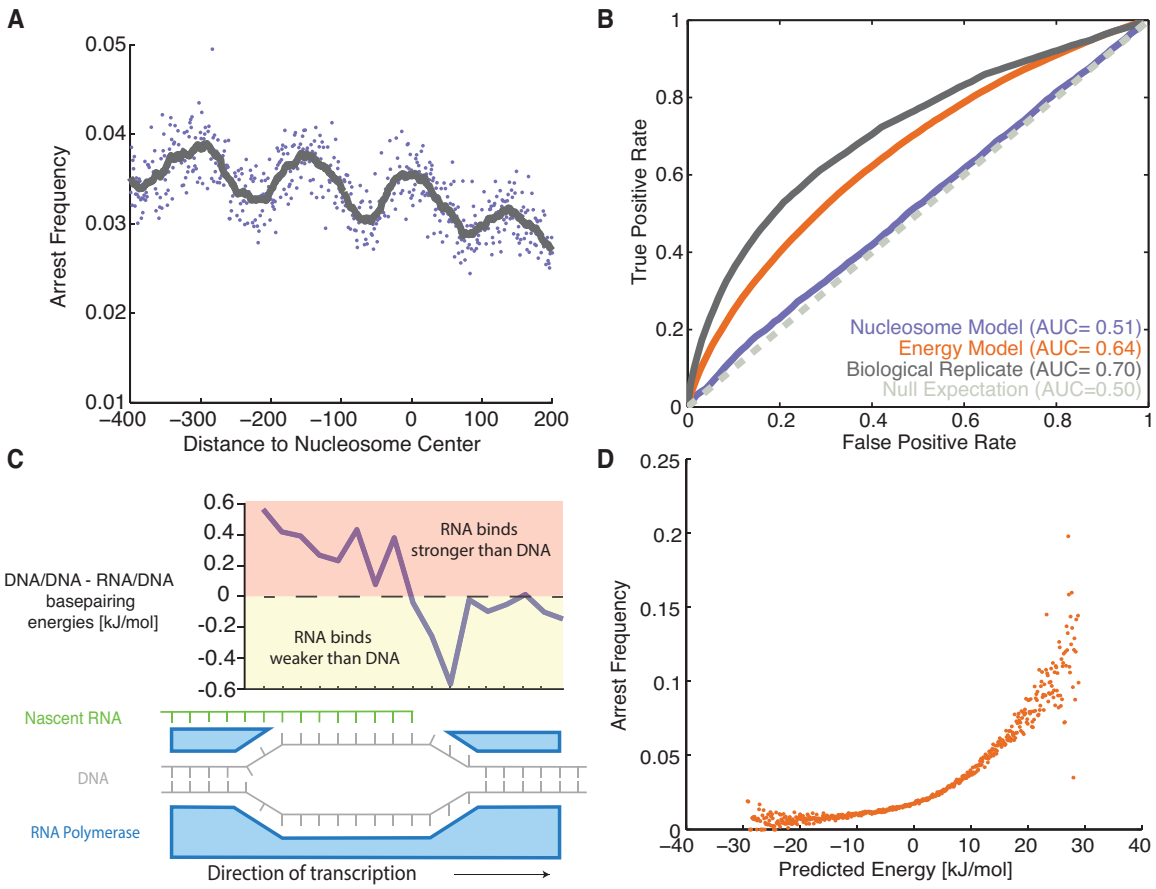

Figure 4.2: Positions of initially arrested RNAP are explained by RNA:DNA and DNA:DNA base pairing more than nucleosome roadblocks. (**A**) Frequency of initial arrests (NET-seq in *dst1*Δ) of RNAP as a function of distance from annotated nucleosomes. Dots represent single base-pairs and lines represent a 25 base-pair running average. Only nucleosomes in coding regions with an average of > 1 read/bp were considered. (**B**) Receiver operating characteristic curves for different proposed models of initial RNAP arrest. The nucleosome classifier was taken to be the distance to the nearest nucleosome (limited to 250 bp away) and only regions with both annotated nucleosomes and a read density > 2 were considered. The energy model was constructed as detailed in the main text. (**C**) Model for initially arrested RNAP in (*dst1*Δ) strains. For all positions annotated as paused, the the average RNA:DNA base-pairing energy minus the DNA:DNA base-pairing energy is shown lined up against the assumed position of the transcriptional bubble. (**D**) The frequency of initial arrests (NET-seq in *dst1*Δ) as a function of the energies predicted from the model in (C).

Other positions that have sequence information are in the RNA:DNA hybrid region of RNAP (Fig. 4.3A). When RNAP is backtracking, RNA:DNA bonds at the downstream end of the RNA:DNA hybrid are being broken, so we hypothesized that breaking the RNA:DNA bonds at the -2 and -3 locations in the transcription bubble might also present a significant barrier to further backtracking after the RNA cleavage. We then used the known RNA:DNA sequence dependent energies to compute the energetics of such a process at each location in the genome.

We combined the contributions from tyrosine stacking with RNA:DNA base pairing at the -2 and -3 locations to devise a metric to predict backtracked polymerases at each position in the genome. Surprisingly, this calculation– which only requires the identity of 3 bases around each position in the genome– was able to predict polymerase pause sites with reasonably high accuracy (AUC=0.64) when compared with the WT replicate AUC of 0.84.

To improve our model we noted that wild-type pauses occur after the initial pausing of RNAP and then subsequent backtracking and clipping by TFIIS. We reasoned that including information about the initial site of pausing as indicated by *dst1*Δ NET-seq density would then allow greater accuracy in determining the sites of backtracked polymerases. We saw that the extent of backtracking is mostly between zero and 2 base pairs preceding the initial pause, with zero backtracking being the most common outcome (Fig. 4.1 C). We therefore augmented the energetic prediction of WT pauses by multiplying the energy predictor at each location by the sum of the *dst1*Δ NET-seq reads zero to two bases ahead, appropriately weighted by the cross correlation of WT and *dst1*Δ NET-seq reads

shown in Figure 4.1 C. This correction effectively gives the number of polymerases that will come across this particular sequence during backtracking. The combination of backtracking information and our energy model predicts the location of backtracked polymerases with an AUC of 0.77. By itself, backtracking information solely from *dst1*Δ reads predicts the location of backtracked polymerases with an AUC of 0.68. This implies that the initial site of pausing and the local nature of backtracking play an important part in where polymerases end after backtracking.

Thus, we have created a plausible biophysical theory for the rate limiting step of RNA polymerase after 3' cleavage by TFIIS. Most of the predictive power of the theory depends on the interaction between Y769 of Rpb2 and the 3' end of the nascent RNA and the fact that RNAP only backtracks to a limited extent after initially pausing. However, there could be other reasons for this pattern [29, 104, 114], and it may be that other residues actually interact with the 3' end of the RNA to produce a preference for adenine at the +1 position. Crystal structures of backtracked RNAP with different templates and NET-seq of mutant RNAP would greatly contribute to this understanding.

4.7 Experimental verification of model predictions

To verify that the biophysical model of RNA polymerase backtracking (WT NET-seq) was correct, we changed the coding regions of two non-essential but highly transcribed genes in order to change the pausing patterns of RNAP: *PHO84* and *BAP2*. Deleting these genes does not cause large fitness defects in rich media [115]. We recoded a 300 base pair region in the middle of each gene (the 700-1000

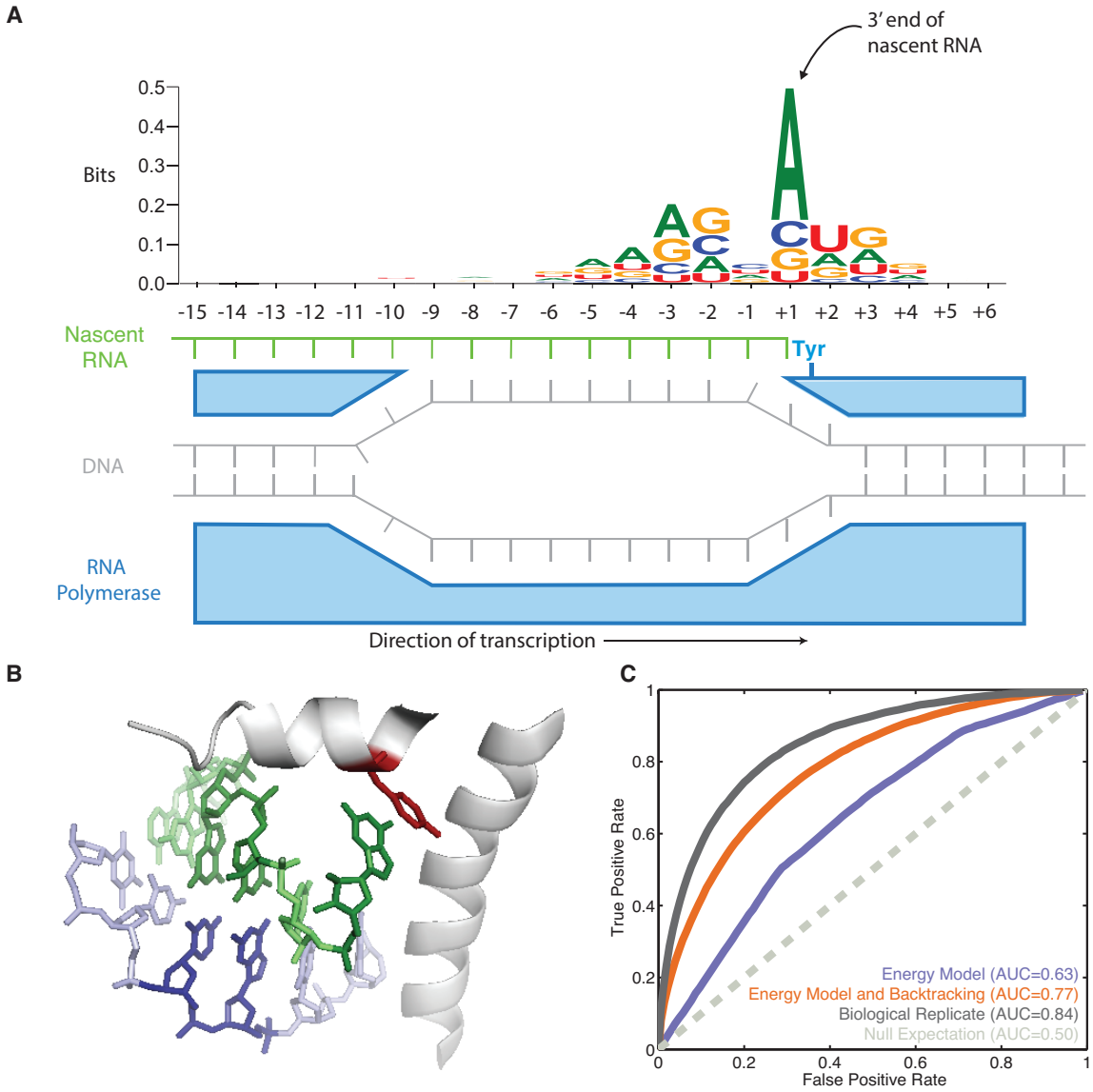

Figure 4.3: Positions of backtracked RNAP can be explained by RNA interactions with RNAP interactions and RNA:DNA base-pairing. (**A**) A sequence weblogo for paused positions for WT NET-seq data is shown aligned with a proposed arrangement of the post-cleavage transcription bubble. The most sequence information is contained at positions +1, -2 and -3. (**B**) Crystal structure of 3' end of nascent mRNA stacked against "gating" tyrosine (Y769 of Rpb2, in red) of RNAP. DNA is in purple, and RNA is in Green. The structure is 3HOZ from [29]. (**C**) Receiver operating characteristic curves and AUCs for the energy model (tyrosine interaction and RNA:DNA pairing), the energy model plus backtracking, and a biological replicate of the WT NET-seq experiment.

base pairs after the start codon) so as not to interfere with polymerase initiation or termination. To control against possible changes in polymerase density from the production of non-functional protein, we also chose to make only synonymous codon changes. These changes were made so as to produce the maximal pausing energy as predicted from the energy metric used above for post-cleavage pausing (Figure 4.4 A). On average the magnitude of the change at each position was about 1 kJ/mol or about the strength of a weak hydrogen bond and less than the thermal fluctuations at 30°C.

Our collaborators in the Churchman lab performed NET-seq on these recoded strains, but we found no systematic increase or decrease in the number of reads in the recoded regions. This may be explained by the fact that coding changes will modify both the initial pausing and subsequent backtracking actions of the polymerase. Thus, even though we designed the changes to increase pausing after backtracking, we would expect both increased and decreased pausing in this region. To check if this was the case, we computed the absolute fold change of NET-seq reads between the recoded and original strains across the entire genome to quantify the extent that the recoding stood out from the un-recoded portions. We then averaged these changes across 300 base-pair segments and normalized them. We found that in both strains, the synonymously recoded regions had the highest absolute fold change across the genome (Figure 4.4 B), meaning that synonymous recoding does in fact cause differences in RNAP density along the recoded region but that the effects are so small that they must be aggregated across the entire recoded region for their effect to be seen. However, these changes are difficult to

precisely interpret and provide only circumstantial support for the specific biophysical model of RNAP pausing we put forward above. The least charitable view of these results are that they simply reflect the difference in capture, amplification and sequencing of two different RNA sequences. In this case the difference in polymerase density due to recoding the genes would be largely overshadowed by sequence-dependent artifacts in the sequencing process.

4.8 Discussion

In this study we have used simple biophysical interactions between RNA, DNA and RNA polymerase to predict and understand RNAP pausing and backtracking. We speculate the the ubiquitous pausing observed in yeast transcription occurs because RNAP has a low threshold for detecting mis-incorporated RNA nucleotides. Sequences where elongation is slightly energetically unfavorable are more likely to trigger pausing and backtracking.

While there are likely other causes for RNAP dynamics, our biophysical model of RNAP pausing could be used to set a baseline when looking for effects from other forms of regulation and may be useful for synthetic biology where certain sequences may be employed or avoided to cause more or fewer transcriptional pauses.

One of the key tests for our biophysical model of RNAP backtracking (WT NET-seq) would be a mutant strain where the "gating" tyrosine-769 of Rpb2p is replaced by another residue. Our collaborators in the labs of Stirling Churchman and Craig Kaplan (UT Austin) were successful in replacing this tyrosine with alanine and phenalanine. However, the NET-seq data from this set of experiments

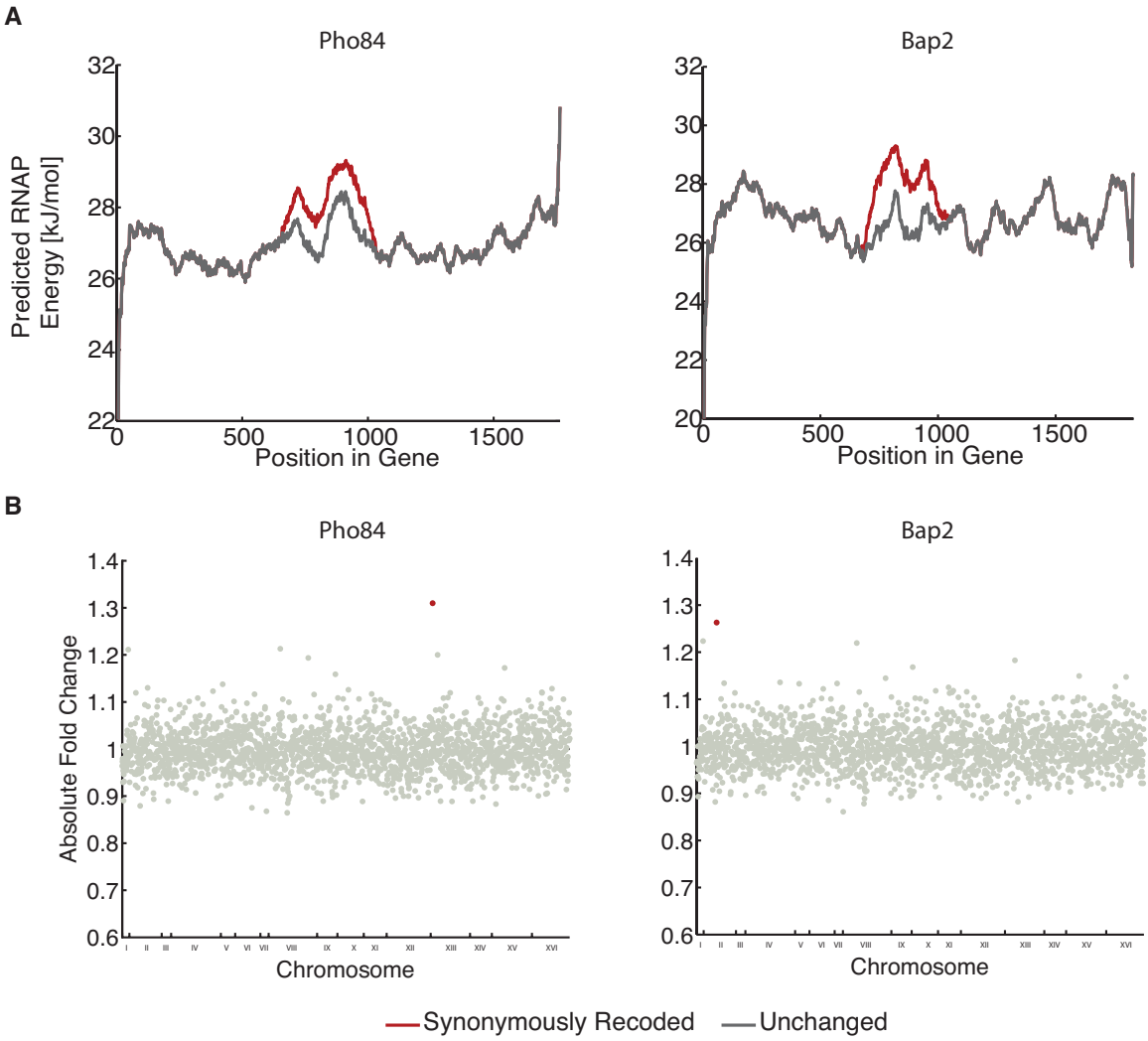

Figure 4.4: Synonymous changes in DNA sequence change RNAP density (**A**). Interaction energy between coding sequence of two genes (Pho84 and Bap2) and RNAP in energy model before (grey) and after being synonymously recoded (red). Energies are smoothed over 100 base pairs. (**B**) Comparison between unchanged (grey) and recoded (red) NET-seq reads across the entire yeast genome. The absolute fold change between reads in all regions were computed for each 300 BP coding region. The average change was normalized to one. The recoded region is shown in red for each strain.

could not be interpreted because of replicability issues.

A hint about the general nature of such biophysical transcription models emerged in the recently developed technique of bacterial NET-seq [116, 117]. It has been shown that bacterial RNAP also has consensus sequence for pausing near the translation start site that may help properly fold RNA and load ribosomes [116]. Changing a specific residue in RNAP changes where it pauses [117], and recoding genes to match the consensus pause sequence causes greater pausing [116]. Avoiding collisions with the ribosome is not likely the case for pausing in eukaryotes like yeast where transcription and translation are performed in separate cellular compartments.

4.9 Materials and methods

4.9.1 Sequence alignment

NET-seq reads from [102] were downloaded from http://www.ncbi.nlm.nih.gov/geo/ via GEO accession number GSE25107. Both this data and newly acquired data was then aligned against the SacCer3 yeast genome database (from UCSC) using Tophat2 [118] and a custom python script created by Julia di Iulio (Churchman lab). This script allowed up to 3 mismatches per alignment and only used alignments that matched uniquely to the yeast genome. Custom Matlab scripts were then written to overlay positions of genes (annotations from SacCer3). The weblogo was generated via http://weblogo.threeplusone.com/ using the background GC frequency for yeast.

Nucleosome positions were downloaded from Weiner et al. [106] and aligned to

the SacCer3 database with a custom matlab script.

4.9.2 Calculation of energy predictions

DNA:DNA binding energies were calculated using a nearest neighbor model with values and procedures detailed in [107]. RNA:DNA energies were similarly calculated using values from [108]. The temperature assumed for free energy calculations was 303 K. RNA:tyrosine stacking energies were taken from [119].

4.9.3 Cloning recoded genes

I recoded sections of *Pho84* and *Bap2* by using the strategy involving the pCORE plasmid outlined in [120]. Briefly, a URA3/KANMX dual cassette was transformed into the region of the genome to be recoded. A 500 base pair oligo (gBlock, IDT) was then synthesized to have homology to the ends of the insertion region and replace the URA3, KANMX dual cassette with the recoded region. Selection was done on 5-FOA to select the replacement. In addition, strains were checked to be G418 sensitive and sequenced to ensure that the correct changes had been made

First, primers RJNS-001 (5'- CTT AGT ATT CAT TCA TTT CTT CGG TGT AAA AGC CTA TGG TGA AAC GGA ATT CAT CTT CAA GAG CTC GTT TTC GAC ACT GG -3') and RJNS-002 (5'- GTC ATC ATT ATA TGG AAC ATT GAA TCC AAT GAG GAT CAT AGT CAA AAG ATA AAT CAC AAC TCC TTA CCA TTA AGT TGA TC -3') for *Bap2* changes and RJNS-006 (5'- TAT CAT CGC TCT TAT CTT GGT TGC TGC TTA CAA GGG CGA ACT AGA ATA CGC AAA CTC TGG GAG CTC GTT TTC GAC ACT GG -3') and RJNS-007

(5'- ACC TAG CAA AAT CTT ACC GTA CTT CCA TTG ACC AAA ATG TCT GCA GAA ATC TTT GAA CGA TCC TTA CCA TTA AGT TGA TC -3') for *Pho84* changes were used to PCR the pCORE plasmid region consisting of the URA3 gene and KANMX resistance cassette. These PCR products were transformed independently in to the WT NET-seq strain YSC001[102] (RJY030) and selected for URA3 auxotraphy and G418 resistance. In addition colony PCR with RJNS-020 (5'- CCA TGA AGT CGC GCC ATG TTG TCA TG -3') and RJNS-021 (5'- CCT TGT ATC CCA ACT CGT CTA GAG ACC -3') for BAP2 or RJNS-022 (5'- CGG TTG GCA ACA AGT TAA GAC CAT CTC C -3') and RJNS-023 (5'- GGA ACA ATA AAG GTG GTT GTG TTT GGA CC -3') were performed to check for insertion in the correct location.

To swap out the markers, gBlocks were ordered for *Bap2*, RJNS-003 (5'- GGG AAT GAT AAA ATA AAT CCG GAC ATT TAT ATT CTT ATT TTC TAT GTT TTC TTA GTA TTC ATT CAT TTC TTC GGT GTA AAA GCC TAT GGT GAA ACG GAA TTC ATC TTC AAC TGC TGC AAA ATA CTG ATG ATA GCG GGA TTC ATC ATA CTA AGC ATA GTA ATC AAC TGC GGG GGA GCC GGG AAC GAC GGG TAC ATA GGG GCA ACC TAC TGG CAT AAC CCA GGG GCA TTC GCA GGG GAC ACA AGC ATA GGT AGG TTC AAA AAC GTA TGC TAC ATA CTG GTA ACA GCT TAC TTC AGC TTC GGG GGA ATG GAA TTA TTC GCG CTA TCG GTA CAG GAG CAA AGC AAC CCA AGA AAG AGC ACG CCG GTA GCA GCC AAG CGA AGC ATA TAC CGA ATA GTT GTG ATT TAT CTT TTG ACT ATG ATC CTC ATT GGA TTC AAT GTT CCA TAT AAT GAT GAC CAA CTA ATG GGC GCA GGC

GGA TCC GCT ACA CAT GCA TCT CCC TAT GTC TT -3') and *Pho84* RJNS-008 (5'- ATG GGT GCT GTC TTT GCT AAC CAA GCT TGG GGT CAA ATC TCC GGT GGT ATC ATC GCT CTT ATC TTG GTT GCT GCT TAC AAG GGC GAA CTA GAA TAC GCA AAC TCT GGA GCG GAA TGC GAC GCA AGA TGC CAA AAG GCA TGC GAC CAA ATG TGG CGG ATA CTA ATC GGA CTG GGG ACG GTA CTG GGA TTG GCA TGT TTG TAC TTC CGG CTA ACA ATT CCG GAA AGC CCC AGG TAC CAA CTA GAT GTG AAC GCA AAG CTA GAA CTG GCA GCA GCG GCA CAA GAG CAA GAC GGG GAA AAG AAA ATT CAC GAC ACA AGC GAC GAA GAC ATG GCA ATA AAC GGA TTG GAG AGG GCA AGC ACG GCA GTG GAA AGC CTG GAC AAC CAC CCG CCA AAG GCA TCG TTC AAA GAT TTC TGC AGA CAT TTT GGT CAA TGG AAG TAC GGT AAG ATT TTG CTA GGT ACT GCT GGT TCA TGG TTT ACC TTA GAT GTT GCT TTC TAC GGG TTG AG -3'). RJNS-003 was amplified further with RJNS-004 (5'- GGG AAT GAT AAA ATA AAT CCG GAC ATT TAT ATT C -3') and RJNS-005 (5'- AAG ACA TAG GGA GAT GCA TGT GTA -3') and transformed into the intermediate strain for *Bap2* recoding to make RJY035. Similarly RJNS-008 was amplified with RJNS-009 (5'- ATG GGT GCT GTC TTT GCT AAC -3') and RJNS-010 (5'- CTC AAC CCG TAG AAA GCA ACA TC -3') for *Pho84* to make RJY036.

After transformation, the strains were checked for 5-FOA resistance and G418 sensitivity and sequenced with RJNS-004 or RJNS-008 appropriately to check that recoding was accomplished.

5

Future Directions

As with all scientific inquiry, the results presented in this thesis bring to light new questions that we were not able to explore fully. This chapter puts forth ideas that future workers may find useful.

In chapter 2, we quantified some of the phenomenology and mechanisms of mitochondrial and mtDNA segregation in yeast, but many questions remain. In

particular, because the observed regular spacing of nucleoids in mitochondria does not fully recapitulate the observed segregation error, other mechanisms that segregate nucleoids could be in play. A particularly intriguing hypothesis is that MMB1p interacts indirectly with mtDNA. MMB1p was shown to help bind mitochondria to microtubules [43], but it is not known whether it binds to certain locations on the mitochondria. If MMB1p interacted with mitochondria at the locations where mtDNA is located, it would provide a direct route for accurate segregation of nucleoids. Simple colocalization studies of MMB1p and mtDNA would test this hypothesis.

There is also a further level of quantification for mtDNA that we were not successful in making: measuring the number of mtDNA genomes per nucleoid. Our staining methods were not able to achieve this level of sensitivity, but the EdU assay could possibly be enhanced to make this possible. However, this assay would be extremely difficult as only a multi-peaked distribution of nucleoid intensities could ensure that the assay really worked (as with mRNA in [121]). A few trials of this type of experiment showed that the cell-to-cell variability in staining or incorporation would have to be taken into account.

For nucleoid replication, our results imply Poisson replication statistics that are immune to both extrinsic noise between cells and do not have any negative feedback control. The mechanism behind this outcome is not understood. It could be that there is actually negative feedback in nucleoid replication but that it is exactly cancelled out by the noise between cells. This is unlikely because of the number of cell division mutants that show very similar replication statistics even though they

likely have very different cell-to-cell variability. More likely, there is a mechanism by which replication becomes insensitive to the number of nucleoids currently present in the cell. Such a system exists in some bacterial plasmid systems where an anti-sense RNA that inhibits replication is produced in direct proportion to the number plasmids in the cell; this results in a Poisson addition of plasmids in each generation [122–124]. No such mechanisms have yet been found in mtDNA, but new mitochondrial gene products are still being discovered [125].

In chapter 3, We presented mathematical arguments that suggest cooperativity may have trade-offs in biological systems trying to achieve many goals to maximize fitness. However, only by actually understanding the molecular details of pattern forming systems, will we know whether these trade-offs are salient. Creation of an *in vivo* synthetic Turing-type system would elucidate the trade-offs when cooperativity is employed, something that has been partially explored [75].

To test whether the actin-based time-delay model we propose is correct, measuring the distribution of times between initial polarization by the diffusion-mediated mechanism and the actin-mediated mechanism would be warrented. A narrow distribution of times would be consistent with our theory and some statistics have already been gathered partially supporting this hypothesis [126].

In chapter 4, we showed how high-resolution, genome-wide data could generate biophysical models of RNA polymerase. Our recoding experiment showed a slight difference in polymerase density in the recoding regions, but perhaps doing the same recoding in a *dst1*Δ strain would also be useful to check where RNAP originally pauses. Recoding the entire yeast genome to allow more or less pausing and

monitoring the paused and backtracked polymerase would be the most comprehensive test of pausing and backtracking; this experiment is beginning to be feasible given technological improvements in genome engineering [127].

In addition, although our collaborators performed the experiment of changing the "gating" tyrosine of RNAP that is thought to interact with the nascent RNA and cause some of the sequence specific pausing, the conclusions were not clear as the controls did not produce the expected results. Redoing this experiment would seem appropriate. It also may be that the exact amount of backtracking of RNAP varies from sample to sample and that this will have to be taken into account to make theory account for the variation in experiments.

Further, although we created an energy model for polymerase density, a more complete model would integrate kinetics in order to more accurately prediction where polymerase backtrack to. For example this would consider not only the energy model for each individual location in the genome but also which energies occur nearby.

References

[1] D. Huh and J. Paulsson, "Random partitioning of molecules at cell division.," *Proceedings of the National Academy of Sciences*, vol. 108, pp. 15004–15009, Sept. 2011.

[2] D. Huh and J. Paulsson, "Non-genetic heterogeneity from stochastic partitioning at cell division," *Nature genetics*, vol. 43, pp. 95–100, Dec. 2010.

[3] M. P. M. Yaffe, D. D. Harata, P. P. Nurse, and 6, "Microtubules mediate mitochondrial distribution in fission yeast.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, pp. 11664–11668, Oct. 1996.

[4] S. M. Rafelski, M. P. Viana, Y. Zhang, Y.-H. M. Chan, K. S. Thorn, P. Yam, J. C. Fung, H. Li, L. d. F. Costa, and W. F. Marshall, "Mitochondrial network size scaling in budding yeast.," *Science*, vol. 338, pp. 822–824, Nov. 2012.

[5] S. Meeusen and J. Nunnari, "Evidence for a two membrane-spanning autonomous mitochondrial DNA replisome," *The Journal of Cell Biology*, vol. 163, pp. 503–510, Nov. 2003.

[6] B. Kornmann, E. Currie, S. R. Collins, M. Schuldiner, J. Nunnari, J. S. Weissman, and P. Walter, "An ER-Mitochondria Tethering Complex Revealed by a Synthetic Biology Screen," *Science*, vol. 325, pp. 477–481, July 2009.

[7] K. Mitra, C. Wunder, B. Roysam, G. Lin, and J. Lippincott-Schwartz, "A hyperfused mitochondrial state achieved at G1–S regulates cyclin E buildup and entry into S phase," *Proceedings of the National Academy of Sciences*, vol. 106, no. 29, p. 11960, 2009.

[8] D. Rapaport, M. Brunner, W. Neupert, and B. Westermann, "Fzo1p is a mitochondrial outer membrane protein essential for the biogenesis of

functional mitochondria in Saccharomyces cerevisiae.," *The Journal of biological chemistry*, vol. 273, pp. 20150–20155, Aug. 1998.

[9] G. J. Capps, D. C. Samuels, and P. F. Chinnery, "A model of the nuclear control of mitochondrial DNA replication.," *Journal of theoretical biology*, vol. 221, pp. 565–583, Apr. 2003.

[10] L. Clay Montier, J. Deng, and Y. Bai, "Number matters: control of mammalian mitochondrial DNA copy number," *Journal of Genetics and Genomics*, vol. 36, no. 3, pp. 125–131, 2009.

[11] M. Satoh and T. Kuroiwa, "Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell," *Experimental Cell Research*, vol. 196, pp. 137–140, Sept. 1991.

[12] E. B. Wilson, "The Distribution of the Chondriosomes to the Spermatozoa in Scorpions.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 2, pp. 321–324, June 1916.

[13] J. Nunnari, W. F. Marshall, A. Straight, A. Murray, J. W. Sedat, and P. Walter, "Mitochondrial transmission during mating in Saccharomyces cerevisiae is determined by mitochondrial fusion and fission and the intramitochondrial segregation of mitochondrial DNA," *Molecular biology of the cell*, vol. 8, pp. 1233–1242, July 1997.

[14] P. Haffter and T. D. Fox, "Nuclear mutations in the petite-negative yeast Schizosaccharomyces pombe allow growth of cells lacking mitochondrial DNA," *Genetics*, vol. 131, no. 2, p. 255, 1992.

[15] H. Hua and S. E. Kearsey, "Monitoring DNA replication in fission yeast by incorporation of 5-ethynyl-2'-deoxyuridine," *Nucleic Acids Research*, vol. 39, pp. e60–e60, May 2011.

[16] D. Bogenhagen and D. Clayton, "Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle," *Cell*, vol. 11, no. 4, pp. 719–727, 1977.

[17] A. Turing, "The chemical basis of morphogenesis," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 237, no. 641, p. 37, 1952.

[18] A. Gierer and H. Meinhardt, "A theory of biological pattern formation," *Kybernetik*, vol. 12, pp. 30–39, Dec. 1972.

[19] C. Chou, Q. Nie, and T. Yi, "Modeling robustness tradeoffs in yeast cell polarization induced by spatial gradients," *PLoS ONE*, vol. 3, no. 9, p. e3103, 2008.

[20] A. Jilkine and L. Edelstein-Keshet, "A Comparison of Mathematical Models for Polarization of Single Eukaryotic Cells in Response to Guided Cues," *PLoS Computational Biology*, vol. 7, p. e1001121, Apr. 2011.

[21] M. Dangkulwanich, T. Ishibashi, S. Liu, M. L. Kireeva, L. Lubkowska, M. Kashlev, and C. J. Bustamante, "Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism," *eLife*, vol. 2, pp. e00971–e00971, Jan. 2013.

[22] E. A. Galburt, S. W. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, and C. Bustamante, "Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner," *Nature*, vol. 446, pp. 820–823, Mar. 2007.

[23] L. Bintu, T. Ishibashi, M. Dangkulwanich, Y.-Y. Wu, L. Lubkowska, M. Kashlev, and C. Bustamante, "Nucleosomal Elementsthat Control the Topography of the Barrier to Transcription," *Cell*, vol. 151, pp. 738–749, Nov. 2012.

[24] L. J. Core, J. J. Waterfall, and J. T. Lis, "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.," *Science*, vol. 322, pp. 1845–1848, Dec. 2008.

[25] H. Kwak, N. J. Fuda, L. J. Core, and J. T. Lis, "Precise maps of RNA polymerase reveal how promoters direct initiation and pausing.," *Science*, vol. 339, pp. 950–953, Feb. 2013.

[26] L. S. Churchman and J. S. Weissman, "Nascent transcript sequencing visualizes transcription at nucleotide resolution," *Nature*, vol. 469, pp. 368–373, Jan. 2011.

[27] L. Bai, R. Fulbright, and M. Wang, "Mechanochemical Kinetics of Transcription Elongation," *Physical Review Letters*, vol. 98, p. 068103, Feb. 2007.

[28] L. Bai, A. Shundrovsky, and M. D. Wang, "Sequence-dependent Kinetic Model for Transcription Elongation by RNA Polymerase," *Journal of Molecular Biology*, vol. 344, pp. 335–349, Nov. 2004.

[29] J. F. Sydow, F. Brueckner, A. C. M. Cheung, G. E. Damsma, S. Dengl, E. Lehmann, D. Vassylyev, and P. Cramer, "Structural Basis of Transcription: Mismatch-Specific Fidelity Mechanismsand Paused RNA Polymerase II with Frayed RNA," *Molecular Cell*, vol. 34, pp. 710–721, June 2009.

[30] E. Nudler, "RNA Polymerase Active Center: The Molecular Engine of Transcription," *Annual review of biochemistry*, vol. 78, pp. 335–361, June 2009.

[31] D. Ó. Maoiléidigh, V. R. Tadigotla, E. Nudler, and A. E. Ruckenstein, "A Unified Model of Transcription Elongation: What Have We Learned from Single-Molecule Experiments?," *Biophysj*, vol. 100, pp. 1157–1166, Mar. 2011.

[32] I. Toulokhonov, J. Zhang, M. Palangat, and R. Landick, "A Central Role of the RNA Polymerase Trigger Loop in Active-Site Rearrangement during Transcriptional Pausing," *Molecular Cell*, vol. 27, pp. 406–419, Aug. 2007.

[33] Z. Liu and R. A. Butow, "Mitochondrial Retrograde Signaling," *Annual review of genetics*, vol. 40, pp. 159–185, Dec. 2006.

[34] S. Hoppins and J. Nunnari, "Mitochondrial Dynamics and Apoptosis–the ER Connection," *Science*, vol. 337, pp. 1052–1054, Aug. 2012.

[35] M. R. Duchen, "Mitochondria in health and disease: perspectives on a new mitochondrial biology," *Molecular Aspects of Medicine*, vol. 25, pp. 365–451, Aug. 2004.

[36] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz, "Gene regulation at the single-cell level.," *Science*, vol. 307, pp. 1962–1965, Mar. 2005.

[37] I. Lestas, G. Vinnicombe, and J. Paulsson, "Fundamental limits on the suppression of molecular fluctuations," *Nature*, vol. 467, pp. 174–178, Sept. 2010.

[38] J. Paulsson, "Summing up the noise in gene networks.," *Nature*, vol. 427, pp. 415–418, Jan. 2004.

[39] M. P. Yaffe, N. Stuurman, and R. D. Vale, "Mitochondrial positioning in fission yeast is driven by association with dynamic microtubules and mitotic spindle poles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, pp. 11424–11428, Sept. 2003.

[40] N. Krüger and I. M. Tolić-Nørrelykke, "Association of mitochondria with spindle poles facilitates spindle alignment," *Current Biology*, vol. 18, no. 15, pp. R646–R647, 2008.

[41] M. P. Viana, S. Lim, and S. M. Rafelski, *Quantifying mitochondrial content in living cells*. Biophysical Methods in Cell Biology, Elsevier Ltd, Jan. 2015.

[42] S. Sivakumar, M. Porter-Goff, P. K. Patel, K. Benoit, and N. Rhind, "In vivo labeling of fission yeast DNA with thymidine and thymidine analogs.," *Methods*, vol. 33, pp. 213–219, July 2004.

[43] C. Fu, D. Jain, J. Costa, G. Velve-Casquillas, and P. T. Tran, "mmb1p Binds Mitochondria to Dynamic Microtubules," *Current Biology*, vol. 21, pp. 1431–1439, Sept. 2011.

[44] M. P. Yaffe, D. Harata, F. Verde, M. Eddison, T. Toda, and P. Nurse, "Microtubules mediate mitochondrial distribution in fission yeast," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, pp. 11664–11668, Oct. 1996.

[45] A. Salic and T. J. Mitchison, "A chemical method for fast and sensitive detection of DNA synthesis in vivo," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, pp. 2415–2420, Feb. 2008.

[46] J. Tauber, A. Dlasková, J. Šantorová, K. Smolková, L. Alán, T. Špaček, L. Plecitá-Hlavatá, M. Jabůrek, and P. Ježek, "The International Journal of Biochemistry & Cell Biology," *International Journal of Biochemistry and Cell Biology*, vol. 45, pp. 593–603, Mar. 2013.

[47] F. Iborra, H. Kimura, and P. Cook, "The functional organization of mitochondrial genomes in human cells," *BMC biology*, vol. 2, no. 1, p. 9, 2004.

[48] D. F. Savage, B. Afonso, A. H. Chen, and P. A. Silver, "Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery," *Science*, vol. 327, pp. 1258–1261, Mar. 2010.

[49] T. Tatsuta, M. Scharwey, and T. Langer, "Mitochondrial lipid trafficking," *Trends in cell biology*, vol. 24, pp. 44–52, Jan. 2014.

[50] A. Murley, L. L. Lackner, C. Osman, M. West, G. K. Voeltz, P. Walter, and J. Nunnari, "ER-associated mitochondrial division links the distribution of mitochondria and mitochondrial DNA in yeast," *eLife*, vol. 2, pp. e00422–e00422, Jan. 2013.

[51] A. Hilfinger and J. Paulsson, "Separating intrinsic from extrinsic fluctuations in dynamic biological systems.," *Proceedings of the National Academy of Sciences*, vol. 108, pp. 12167–12172, July 2011.

[52] M. Elowitz, A. Levine, E. Siggia, and P. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, no. 5584, p. 1183, 2002.

[53] Z. Chu, J. Li, M. Eshaghi, R. K. Karuturi, K. Lin, and J. Liu, "Adaptive expression responses in the Pol-γ null strain of S. pombe depleted of mitochondrial genome," *BMC Genomics*, vol. 8, no. 1, p. 323, 2007.

[54] S. O. Skinner, L. A. S. u. lveda, H. Xu, and I. Golding, "Measuring mRNA copy number in individual Escherichia coli cells using single-moleculefluorescent in situ hybridization," *Nature Protocols*, vol. 8, pp. 1100–1113, May 2013.

[55] O. Sliusarenko, J. Heinritz, T. Emonet, and C. Jacobs-Wagner, "High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics," *Molecular Microbiology*, vol. 80, pp. 612–627, Mar. 2011.

[56] F. R. Neumann and P. Nurse, "Nuclear size control in fission yeast," *The Journal of Cell Biology*, vol. 179, pp. 593–600, Nov. 2007.

[57] J. Q. Wu, "Counting Cytokinesis Proteins Globally and Locally in Fission Yeast," *Science*, vol. 310, pp. 310–314, Oct. 2005.

[58] D.-U. Kim, J. Hayles, D. Kim, V. Wood, H.-O. Park, M. Won, H.-S. Yoo, T. Duhig, M. Nam, G. Palmer, S. Han, L. Jeffery, S.-T. Baek, H. Lee, Y. S. Shim, M. Lee, L. Kim, K.-S. Heo, E. J. Noh, A.-R. Lee, Y.-J. Jang, K.-S. Chung, S.-J. Choi, J.-Y. Park, Y. Park, H. M. Kim, S.-K. Park, H.-J. Park, E.-J. Kang, H. B. Kim, H.-S. Kang, H.-M. Park, K. Kim, K. Song, K. B. Song, P. Nurse, and K.-L. Hoe, "Analysis of a genome-wide set of gene

deletions in the fission yeast Schizosaccharomyces pombe," *Nature Biotechnology*, vol. 28, pp. 617–623, May 2010.

[59] D. G. Gibson, G. A. Benders, K. C. Axelrod, J. Zaveri, M. A. Algire, M. Moodie, M. G. Montague, J. C. Venter, H. O. Smith, and C. A. Hutchison, "One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome.," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 20404–20409, Dec. 2008.

[60] B. Westermann and W. Neupert, "Mitochondria-targeted green fluorescent proteins: convenient tools for the study of organelle biogenesis in Saccharomyces cerevisiae," *Yeast (Chichester, England)*, vol. 16, no. 15, pp. 1421–1427, 2000.

[61] J. Bähler, J. Q. Wu, M. S. Longtine, N. G. Shah, A. McKenzie, A. B. Steever, A. Wach, P. Philippsen, and J. R. Pringle, "Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe.," *Yeast (Chichester, England)*, vol. 14, pp. 943–951, July 1998.

[62] S. Sick, S. Reinker, J. Timmer, and T. Schlake, "WNT and DKK determine hair follicle spacing through a reaction-diffusion mechanism.," *Science*, vol. 314, pp. 1447–1450, Dec. 2006.

[63] J. Raspopovic, L. Marcon, L. Russo, and J. Sharpe, "Modeling digits. Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients.," *Science*, vol. 345, pp. 566–570, Aug. 2014.

[64] A. B. Goryachev and A. V. Pokhilko, "Dynamics of Cdc42 network embodies a Turing-type mechanism of yeast cell polarity," *FEBS Letters*, vol. 582, pp. 1437–1443, Apr. 2008.

[65] T. W. Hiscock and S. G. Megason, "Mathematically guided approaches to distinguish models of periodic patterning," *Development*, vol. 142, pp. 409–419, Jan. 2015.

[66] A. R. Houk, A. Jilkine, C. O. Mejean, R. Boltyanskiy, E. R. Dufresne, S. B. Angenent, S. J. Altschuler, L. F. Wu, and O. D. Weiner, "Membrane Tension Maintains Cell Polarity by Confining Signals to the Leading Edge during Neutrophil Migration," *Cell*, vol. 148, pp. 175–188, Jan. 2012.

[67] A. Bender and J. Pringle, "Multicopy suppression of the cdc24 budding defect in yeast by CDC42 and three newly identified genes including the ras-related gene RSR1," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 86, no. 24, p. 9976, 1989.

[68] A. Butty, N. Perrinjaquet, A. Petit, M. Jaquenoud, J. Segall, K. Hofmann, C. Zwahlen, and M. Peter, "A positive feedback loop stabilizes the guanine-nucleotide exchange factor Cdc24 at sites of polarization," *The EMBO Journal*, vol. 21, no. 7, pp. 1565–1576, 2002.

[69] I. Sagot, S. K. Klee, and D. Pellman, "Yeast formins regulate cell polarity by controlling the assembly of actin cables," *Nature Cell Biology*, vol. 4, no. 1, pp. 42–50, 2002.

[70] A. Layton, N. Savage, A. Howell, and S. Carroll, "Modeling Vesicle Traffic Reveals Unexpected Consequences for Cdc42p-Mediated Polarity Establishment," *Current Biology*, 2011.

[71] S. Etienne-Manneville, "Cdc42–the centre of polarity," *Journal of Cell Science*, vol. 117, pp. 1291–1300, Mar. 2004.

[72] S. Etienne-Manneville and A. Hall, "Rho GTPases in cell biology," *Nature*, vol. 420, pp. 629–635, Dec. 2002.

[73] A. Hunding and R. Engelhardt, "Early biological morphogenesis and nonlinear dynamics," *Journal of theoretical biology*, vol. 173, no. 4, pp. 401–413, 1995.

[74] L. Diambra, V. R. Senthivel, D. B. Menendez, and M. Isalan, "Cooperativity To Increase Turing Pattern Space for Synthetic Biology," *ACS Synthetic Biology*, p. 140822101835005, Aug. 2014.

[75] A. H. Chau, J. M. Walter, J. Gerardin, C. Tang, and W. A. Lim, "Designing Synthetic Regulatory Networks Capable of Self-Organizing Cell Polarization," *Cell*, vol. 151, pp. 320–332, Oct. 2012.

[76] J. E. Irazoqui, "Polarity establishment in yeast," *Journal of Cell Science*, vol. 117, pp. 2169–2171, May 2004.

[77] E. Marco, R. Wedlich-Soldner, R. Li, S. J. Altschuler, and L. F. Wu, "Endocytosis optimizes the dynamic localization of membrane proteins that regulate cortical polarity," *Cell*, vol. 129, pp. 411–422, Apr. 2007.

[78] A. Mogilner, J. Allard, and R. Wollman, "Cell Polarity: Quantitative Modeling as a Tool in Cell Biology," *Science*, vol. 336, pp. 175–179, Apr. 2012.

[79] S. J. Altschuler, S. B. Angenent, Y. Wang, and L. F. Wu, "On the spontaneous emergence of cell polarity," *Nature*, vol. 454, pp. 886–889, Aug. 2008.

[80] R. Wedlich-Soldner, S. C. Wai, T. Schmidt, and R. Li, "Robust cell polarity is a dynamic state established by coupling transport and GTPase signaling," *The Journal of Cell Biology*, vol. 166, pp. 889–900, Sept. 2004.

[81] D. Johnson, "Cdc42: an essential Rho-type GTPase controlling eukaryotic cell polarity," *Microbiology and Molecular Biology Reviews*, vol. 63, no. 1, p. 54, 1999.

[82] O. Brandman, J. E. Ferrell, R. Li, and T. Meyer, "Interlinked fast and slow positive feedback loops drive reliable cell decisions," *Science*, vol. 310, pp. 496–498, Oct. 2005.

[83] A. Jilkine, S. B. Angenent, L. F. Wu, and S. J. Altschuler, "A density-dependent switch drives stochastic clustering and polarization of signaling molecules.," *PLoS Computational Biology*, vol. 7, p. e1002271, Nov. 2011.

[84] A. Howell, N. Savage, S. Johnson, I. Bose, A. Wagner, T. Zyla, H. Nijhout, M. Reed, A. Goryachev, and D. Lew, "Singularity in polarization: rewiring yeast cells to make two buds," *Cell*, vol. 139, no. 4, pp. 731–743, 2009.

[85] A. Jilkine, S. B. Angenent, L. F. Wu, and S. J. Altschuler, "A Density-Dependent Switch Drives Stochastic Clustering and Polarization of Signaling Molecules," *PLoS Computational Biology*, vol. 7, p. e1002271, Nov. 2011.

[86] E. Marco, R. Wedlich-Soldner, R. Li, S. Altschuler, and L. Wu, "Endocytosis optimizes the dynamic localization of membrane proteins that regulate cortical polarity," *Cell*, vol. 129, no. 2, pp. 411–422, 2007.

[87] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer, "BioNumbers–the database of key numbers in molecular and cell biology," *Nucleic Acids Research*, vol. 38, pp. D750–D753, Dec. 2009.

[88] A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, and R. Milo, "The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters," *Biochemistry*, vol. 50, pp. 4402–4410, May 2011.

[89] B. D. Slaughter, A. Das, J. W. Schwartz, B. Rubinstein, and R. Li, "Dual modes of cdc42 recycling fine-tune polarized morphogenesis," *Developmental Cell*, vol. 17, pp. 823–835, Dec. 2009.

[90] R. Wedlich-Soldner, S. Altschuler, L. Wu, and R. Li, "Spontaneous cell polarization through actomyosin-based delivery of the Cdc42 GTPase," *Science*, vol. 299, pp. 1231–1235, Feb. 2003.

[91] J. Irazoqui, A. Howell, C. Theesfeld, and D. Lew, "Opposing roles for actin in Cdc42p polarization," *Molecular biology of the cell*, vol. 16, no. 3, p. 1296, 2005.

[92] J. Pedraza and J. Paulsson, "Effects of molecular memory and bursting on fluctuations in gene expression," *Science*, vol. 319, no. 5861, p. 339, 2008.

[93] Å. Engqvist-Goldstein and D. Drubin, "ACTIN ASSEMBLY AND ENDOCYTOSIS: From Yeast to Mammals," *Annual Review of Cell and Developmental Biology*, vol. 19, no. 1, pp. 287–332, 2003.

[94] A. V. Hill, "The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves.," *Proceedings of the Physiological Society*, pp. 1–7, Jan. 1910.

[95] K. Adelman and J. T. Lis, "Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans," *Nature Reviews Genetics*, vol. 13, pp. 720–731, Sept. 2012.

[96] S. Nechaev, D. C. Fargo, G. dos Santos, L. Liu, Y. Gao, and K. Adelman, "Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in Drosophila," *Science*, vol. 327, pp. 335–338, Jan. 2010.

[97] T. Henriques, D. A. Gilchrist, S. Nechaev, M. Bern, G. W. Muse, A. Burkholder, D. C. Fargo, and K. Adelman, "Stable Pausing by RNA Polymerase II Provides an Opportunity to Target and Integrate Regulatory Signals," *Molecular Cell*, vol. 52, pp. 517–528, Nov. 2013.

[98] B. Zamft, L. Bintu, T. Ishibashi, and C. Bustamante, "Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases.," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 8948–8953, June 2012.

[99] Y. X. Mejia, E. Nudler, and C. Bustamante, "Trigger loop folding determines transcription rate of Escherichia coli'sRNA polymerase," *Proceedings of the National Academy of Sciences*, vol. 112, pp. 743–748, Jan. 2015.

[100] Y. X. Mejia, H. Mao, N. R. Forde, and C. Bustamante, "Thermal Probing of E. coli RNA Polymerase Off-Pathway Mechanisms," *Journal of Molecular Biology*, vol. 382, pp. 628–637, Oct. 2008.

[101] M. Palangat and R. Landick, "Roles of RNA:DNA hybrid stability, RNA structure, and active site conformation in pausing by human RNA polymerase II," *Journal of Molecular Biology*, vol. 311, pp. 265–282, Aug. 2001.

[102] L. S. Churchman and J. S. Weissman, "Nascent transcript sequencing visualizes transcription at nucleotide resolution," *Nature*, vol. 469, pp. 368–373, Jan. 2011.

[103] M. Depken, J. M. R. Parrondo, and S. W. Gril, "Irregular transcription dynamics for rapid production of high-fidelity transcripts," *Audio, Transactions of the IRE Professional Group on*, pp. –, Jan. 2012.

[104] A. C. M. Cheung and P. Cramer, "Structural basis of RNA polymerase II backtracking, arrest and reactivation," *Nature*, vol. 471, pp. 249–253, Apr. 2012.

[105] M. Depken, E. A. Galburt, and S. W. Grill, "The Origin of Short Transcriptional Pauses," *Biophysj*, vol. 96, pp. 2189–2193, Mar. 2009.

[106] A. Weiner, A. Hughes, M. Yassour, O. J. Rando, and N. Friedman, "High-resolution nucleosome mapping reveals transcription-dependent promoter packaging," *Genome research*, vol. 20, pp. 90–100, Jan. 2010.

[107] J. SantaLucia, "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, pp. 1460–1465, Feb. 1998.

[108] N. Sugimoto, S. Nakano, M. Katoh, A. Matsumura, H. Nakamuta, T. Ohmichi, M. Yoneyama, and M. Sasaki, "Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes.," *Biochemistry*, vol. 34, pp. 11211–11216, Sept. 1995.

[109] M. Imashimizu, T. Oshima, L. Lubkowska, and M. Kashlev, "Direct assessment of transcription fidelity by high-resolution RNA sequencing," *Nucleic Acids Research*, vol. 41, pp. 9090–9104, Oct. 2013.

[110] V. Svetlov and E. Nudler, "Biochimica et Biophysica Acta," *BBA - Gene Regulatory Mechanisms*, vol. 1829, pp. 20–28, Jan. 2013.

[111] F. W. Martinez-Rucobo and P. Cramer, "Biochimica et Biophysica Acta," *BBA - Gene Regulatory Mechanisms*, vol. 1829, pp. 9–19, Jan. 2013.

[112] L. R. Rutledge, L. S. Campbell-Verduyn, and S. D. Wetmore, "Characterization of the stacking interactions between DNA or RNA nucleobases and the aromatic amino acids," *Chemical Physics Letters*, vol. 444, pp. 167–175, Aug. 2007.

[113] L. R. Rutledge, H. F. Durst, and S. D. Wetmore, "Computational comparison of the stacking interactions between the aromatic amino acids and the natural or (cationic) methylated nucleobases," *Physical Chemistry Chemical Physics*, vol. 10, no. 19, p. 2801, 2008.

[114] A. C. M. Cheung and P. Cramer, "A Movie of RNA Polymerase II Transcription," *Cell*, vol. 149, pp. 1431–1437, June 2012.

[115] D. Breslow, D. Cameron, S. Collins, M. Schuldiner, J. Stewart-Ornstein, H. Newman, S. Braun, H. Madhani, N. Krogan, and J. Weissman, "A comprehensive strategy enabling high-resolution functional analysis of the yeast genome," *Nature Methods*, vol. 5, no. 8, pp. 711–718, 2008.

[116] M. H. Larson, R. A. Mooney, J. M. Peters, T. Windgassen, D. Nayak, C. A. Gross, S. M. Block, W. J. Greenleaf, R. Landick, and J. S. Weissman, "A Pause Sequence Enriched at Translation Start Sites Drives Transcription Dynamics in Vivo," *Science*, May 2014.

[117] I. O. Vvedenskaya, H. Vahedian-Movahed, J. G. Bird, J. G. Knoblauch, S. R. Goldman, Y. Zhang, R. H. Ebright, and B. E. Nickels, "Interactions between RNA polymerase and the "core recognition element" counteract pausing," *Science*, vol. 344, pp. 1285–1289, June 2014.

[118] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg, "TopHat2: accurate alignment of transcriptomes inthe presence of insertions, deletions and genefusions," *Genome Biology*, vol. 14, p. R36, Apr. 2013.

[119] L. R. Rutledge, L. S. Campbell-Verduyn, K. C. Hunter, and S. D. Wetmore, "Characterization of Nucleobase−Amino Acid Stacking Interactions Utilized by a DNA Repair Enzyme," *The Journal of Physical Chemistry B*, vol. 110, pp. 19652–19663, Oct. 2006.

[120] F. Storici and M. A. Resnick, "The Delitto Perfetto Approach to In Vivo Site-Directed Mutagenesis and Chromosome Rearrangements with Synthetic Oligonucleotides in Yeast," pp. 329–345, Elsevier, 2006.

[121] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, "Real-Time Kinetics of Gene Activity in Individual Bacteria," *Cell*, vol. 123, pp. 1025–1036, Dec. 2005.

[122] J. Paulsson, "Multileveled selection on plasmid replication.," *Genetics*, vol. 161, pp. 1373–1384, Aug. 2002.

[123] K. Nordström, "Plasmid R1–Replication and its control," *Plasmid*, vol. 55, no. 1, pp. 1–26, 2006.

[124] S. Brantl, "Antisense RNAs in plasmids: control of replication and maintenance.," *Plasmid*, vol. 48, pp. 165–173, Nov. 2002.

[125] C. Lee, K. Yen, and P. Cohen, "Humanin: a harbinger ofmitochondrial-derived peptides?," *Trends in Endocrinology & Metabolism*, vol. 24, pp. 222–228, May 2013.

[126] A. S. Howell, N. S. Savage, S. A. Johnson, I. Bose, A. W. Wagner, T. R. Zyla, H. F. Nijhout, M. C. Reed, A. B. Goryachev, and D. J. Lew, "Singularity in polarization: rewiring yeast cells to make two buds.," *Cell*, vol. 139, pp. 731–743, Nov. 2009.

[127] N. Annaluru, H. Muller, L. A. Mitchell, S. Ramalingam, G. Stracquadanio, S. M. Richardson, J. S. Dymond, Z. Kuang, L. Z. Scheifele, E. M. Cooper, Y. Cai, K. Zeller, N. Agmon, J. S. Han, M. Hadjithomas, J. Tullman, K. Caravelli, K. Cirelli, Z. Guo, V. London, A. Yeluru, S. Murugan, K. Kandavelou, N. Agier, G. Fischer, K. Yang, J. A. Martin, M. Bilgel, P. Bohutski, K. M. Boulier, B. J. Capaldo, J. Chang, K. Charoen, W. J. Choi,

P. Deng, J. E. DiCarlo, J. Doong, J. Dunn, J. I. Feinberg, C. Fernandez, C. E. Floria, D. Gladowski, P. Hadidi, I. Ishizuka, J. Jabbari, C. Y. L. Lau, P. A. Lee, S. Li, D. Lin, M. E. Linder, J. Ling, J. Liu, J. Liu, M. London, H. Ma, J. Mao, J. E. McDade, A. McMillan, A. M. Moore, W. C. Oh, Y. Ouyang, R. Patel, M. Paul, L. C. Paulsen, J. Qiu, A. Rhee, M. G. Rubashkin, I. Y. Soh, N. E. Sotuyo, V. Srinivas, A. Suarez, A. Wong, R. Wong, W. R. Xie, Y. Xu, A. T. Yu, R. Koszul, J. S. Bader, J. D. Boeke, and S. Chandrasegaran, "Total synthesis of a functional designer eukaryotic chromosome.," *Science*, vol. 344, pp. 55–58, Apr. 2014.